GENERATION AND CHARACTERIZATION OF TRANSCRIPTION ACTIVATOR-LIKE EFFECTOR NUCLEASES

LAURA WALTL

Dissertation zur Erlangung des Akademischen Grades des Doktors der Naturwissenschaften (Dr. rer. nat)

Institute für Biochemie
Fachbereich 08
Justus Liebig Universität
– January 2016 –
This dissertation was prepared at the Institute for Biochemistry in the department of Biology and Chemistry (FB08) of the Justus Liebig University Giessen. The DFG financed graduate college “Enzymes and Multienzyme Complexes acting on Nucleic Acids” (IRTG 1393) sponsored the work.

**Primary Referee:**
Prof. Dr. Peter Friedhoff
Institut für Biochemie
Fachbereich Biologie und Chemie
Heinrich-Buff-Ring 58
35392 Giessen

**Secondary Referee:**
Prof. Dr. Alexander Brehm
Institut für Molekularbiologie und Tumorforschung (IMT)
Philipps-Universität Marburg
Emil-Mannkopff-Str. 2
35033 Marburg

**Supervisors:**
Dr. Wolfgang Wende

IRTG Supervisors:
(I) Prof. Dr. Alfred Pingoud, Giessen
(II) Prof. Dr. Alexander Brehm, Marburg
(III) Prof. Dr. Virginijus Siksnys, Vilnius

Laura Waltl: *Generation and characterization of Transcription activator-like effector nucleases*, Dissertation zur Erlangung des Akademischen Grades des Doktors der Naturwissenschaften (Dr. rer. nat), © January 2016
Erkenne dich selbst!

― Thales von Milet (≈ 625–≈547 v. Chr.)

Dedicated to my parents.
Gert & Angelika Waltl
ABSTRACT

Designer nucleases are developed to modify the genome in various organisms, and are promising molecular tools for the advancement of gene therapy. They should introduce only one specific double-strand or single-strand break in a complex genome in order to trigger the DNA repair mechanisms of the cell such that the DNA damage is repaired by homologous directed repair (HDR) or non-homologous end joining (NHEJ), an error-prone repair mechanism that can be applied to knock out genes. With the addition of an external DNA fragment, HDR can be exploited to correct or add a gene.

Transcription activator-like effector (TALE) nucleases belong to a platform of designer nucleases and are known to be highly specific, efficient and easy to customize for any given target. Nevertheless, there is still a need to improve the accuracy of TALE nucleases for their reliable application in vivo. To this end, off-target activity should be avoided without compromising cleavage efficiency. This can be achieved by optimization of the catalytic domain as well as by improving the binding domain.

The subjects of this comparative study are the homing endonuclease I-SceI, the widely used TALE-FokI fusion protein, the new RNA guided Cas9 nuclease and our recently developed TALE-PvuII variants as well as the TALE-MutH fusion protein, whose nuclease domain is a sequence- and strand-specific nickase. In the study we have analyzed the toxicity and cleavage efficiency of these specific nucleases in a yeast based single-strand annealing assay as well as in vitro. The present study also focus on the characterization of the in vitro binding mechanism of TALE. Binding characteristics of AvrBs3(ΔN152-C28) were analyzed via anisotropy measurements, specific single- and double-labeled TALE variants were designed and DNA::Protein complexes were analyzed employing FRET measurements. We gained labeled variants remaining their binding characteristics and enables the investigation of the in vitro binding mechanism of TALE proteins.

Our results demonstrate convincingly that the monomeric TALE-MutH nickase displays less toxicity in Saccharomyces cerevisiae compared to the widely used dimeric TALE-FokI nuclease and is able to introduce the repair of the DNA damage via single-strand annealing as efficient as I-SceI, the gold standard for cleaving genomic DNA. To improve the design of TALE-MutH nickases, we analyzed the cleavage behavior on various specific loci in vitro and in vivo. The results presented show that the TALE-MutH fusion protein is a promising programmable nickase for in vivo applications and these results were confirmed by a TALE-MutH addressing a relevant therapeutic gene.
ZUSAMMENFASSUNG


CONTENTS

i INTRODUCTION 1

1 INTRODUCTION 3
  1.1 Tools for Genome editing 5
    1.1.1 Meganucleases/Homing endonucleases 5
    1.1.2 Fusion-proteins 5
    1.1.3 Clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease Cas9 10
    1.1.4 Catalytic domains 12
  1.2 Therapeutic applications 17
    1.2.1 Delivery systems 17
    1.2.2 Impact of epigenetics on enzyme activation 19
    1.2.3 Therapeutic potential 19
  1.3 Aim 22

ii MATERIAL AND METHODS 23

2 MATERIAL AND METHODS 25
  2.1 Materials 25
    2.1.1 Chemical 25
    2.1.2 Commercial DNA purification kits 25
    2.1.3 Columns/matrices 25
    2.1.4 Dialyse tube 26
    2.1.5 Buffers and media 26
    2.1.6 Fluorophores 31
    2.1.7 Plasmids 32
    2.1.8 E. coli and Saccharomyces cerevisiae strains 34
    2.1.9 Enzymes and markers 35
    2.1.10 Primer and oligonucleotide 35
  2.2 Methods 38
    2.2.1 Molecular cloning 38
    2.2.2 vectors 41
    2.2.3 Protein expression and purification 45
    2.2.4 Eukaryote cell systems 46
    2.2.5 Cleavage assays 48
    2.2.6 Fluorescence based methods 49

iii RESULTS AND DISCUSSION 55

3 RESULTS 57
  3.1 Characterization of highly specific nucleases in vitro and in vivo 57
    3.1.1 Characterization of TALENs 57
    3.1.2 Characterization of RNA guided Cas9 protein in vivo 72
3.1.3 Alignment search of the addressed sites in the genome of Saccharomyces cerevisiae 74
3.1.4 Addressing X-linked Retinitis Pigmentosa with specific nucleases 76
3.2 Characterization of binding characteristics of AvrBs3 79
   3.2.1 Variants of AvrBs3 79
   3.2.2 Binding character of the AvrBs3 variants 82
   3.2.3 Labeled AvrBs3 variants 88
4 DISCUSSION 101
   4.1 Characterization of highly specific nucleases \textit{in vitro} and \textit{in vivo} 101
      4.1.1 Beyond AvrBs4-MutH 104
   4.2 Characterization of binding characteristics of AvrBs3 109
iv APPENDIX 117
A APPENDIX 119
   A.1 Supplementary experiments 119

BIBLIOGRAPHY 123
LIST OF FIGURES

Figure 1  Gene targeting with meganucleases, fusionproteins or RNA-guided endonucleases  4
Figure 2  Crystal structure of TALE  7
Figure 3  CRISPR-Cas system  11
Figure 4  Cleavage domain of FokI, restriction enzyme PvuII and MutH  13
Figure 5  The outer segment of the photoreceptore cells  21
Figure 6  Chemical structures of fluorophores  31
Figure 7  Chemical structures of fluorophores  32
Figure 8  Basic steps of the cloning strategies  39
Figure 9  X-Gal converted to a blue dye  40
Figure 10 Golden Gate cloning strategy  42
Figure 11 Single strand annealing  46
Figure 12 Thiol Michael-type addition reaction  49
Figure 13 Architecture of AvrBs4-MutH  57
Figure 14 Design of specific single and double targets  58
Figure 15 Kinetic cleavage assay on T-3-H and H-3-T  59
Figure 16 Cleavage assay with labeled PCR fragments  60
Figure 17 In vitro characterization of AvrBs4-MutH on single targets (T-3-H, T-6-H and H-3-T)  61
Figure 18 In vitro characterization of AvrBs4-MutH on double targets (T-3-H-3-T and T-3-HH-3-T).  61
Figure 19 Kinetics of DNA nicking by AvrBs4-MutH on the double targets (T-3-H-3-T, T-3-HH-3-T)  62
Figure 20 Single strand annealing assay  64
Figure 21 Relative survival of yeast colonies  65
Figure 22 β-galactosidase activity of specific nucleases (I-SceI, AvrBs4-FokI, and AvrBs4-MutH) in S. cerevisiae  66
Figure 23 Yeast colonies growing on X-gal containing ma-
trix  67
Figure 24 AvrBs3-PvuII variants  68
Figure 25 Recognition sequence of AvrB3-PvuII variants  69
Figure 26 Relative survival of yeast colonies  70
Figure 27 Relative β-gal activity of specific nucleases (I-SceI and AvrBs3-PvuII variants) in S. cerevisiae as
analyzed by the ratio of blue colonies expressed in percentage [%]  71
Figure 28 β-gal activity of specific nucleases (I-SceI and AvrBs3-PvuII variants) in S. cerevisiae  72
Figure 29 Architecture of Cas9:gRNA and its recognition sequence  73
Figure 30  β-gal activity of Cas9::gRNA in *Saccharomyces cerevisiae*  73

Figure 31  Sequence within the RPGR gene targeted by the designed TALE-MutH variants  76

Figure 32  Analysis of TALE-MutH addressing GATC (I) site within the ORF15 of the RPGR gene  78

Figure 33  Architecture of AvrBs3 variants  79

Figure 34  Single-site mutation of the cysteine residue  80

Figure 35  Purification of AvrBs3 (B)  81

Figure 36  K_D-values of AvrBs3(ΔN152-C28) and AvrBs3 (B)  82

Figure 37  Determination of the active fractions of AvrBs3(ΔN152-C28) and AvrBs3 (B)  83

Figure 38  Individual competition of AvrBs3 (B)  84

Figure 39  Consecutive competition of AvrBs3 (B)  85

Figure 40  EMSA of AvrBS3 (AB)  87

Figure 41  Determination of the K_D-value of AvrBs3 (B)-Alexa 488  88

Figure 42  SDS-PAGE of single labeled AvrBs3 variants with the fluorophore Cy5  89

Figure 43  Emission spectra of titrating Cy5 labeled AvrBs3 to labeled 5’-end Cy3 specific DNA (I)  90

Figure 44  Spectra of AvrBs3-Cy5 to determine the correction factor α  91

Figure 45  FRET distance measurements of AvrBs3 (A)-Cy5 and AvrBs3 (B)-Cy5 on DNA (I)  92

Figure 46  FRET distance measurements of AvrBs3 (A)-Cy5 and AvrBs3 (B)-Cy5 on DNA (II)  93

Figure 47  SDS-PAGE of AvrBs3 (AB) labeled with Alexa 594, Alexa 488, or Alexa 647  96

Figure 48  Merged fluorescent images of AvrBs3 (AB) labeled with single fluorophores (Alexa 594, Alexa 488, or Alexa 647)  97

Figure 49  EMSA with AvrBs3 (AB) labeled with distinct fluorophores (Alexa 594, Alexa 488, and Alexa 647)  98

Figure 50  Double labeling of AvrBs3 (AB) with Alexa 647 and Alexa 488  99

Figure 51  Emission spectra of AvrBs3 (AB)-Alexa 488-Alexa 647  100

Figure 52  Binding characteristics of AvrBs3 (AB) - Alexa 488-Alexa 647  100

Figure 53  “close-state” and “far-state”  114

Figure 54  Emission spectra of titrating Cy5 labeled AvrBs3 (B) to 5’-end Cy3 labeled specific DNA (I)  119
Figure 55  Emission spectra of titrating Cy5 labeled AvrBs3 to 5'-end Cy3 labeled specific DNA (II)  120
Figure 56  Spectra of DNA-Cy3 to determined the correction factor $\beta$  121

LIST OF TABLES

Table 1  Commercial purification kits for DNA  25
Table 2  Columns and matrices  25
Table 3  Buffers for assays  26
Table 4  Buffers for DNA electrophoresis  27
Table 5  Buffers for Saccharomyces cerevisiae  28
Table 6  Media for E. coli and Saccharomyces cerevisiae  29
Table 7  LB-plates and drop out plates  30
Table 8  Characteristics of fluorophores  31
Table 9  Addressed sites  34
Table 10  Maker for gel electrophoresis  35
Table 11  Cassettes used as inserts  35
Table 12  Cassettes used as target DNA  36
Table 13  Primers used to create Megaprimers  36
Table 14  Primers  37
Table 15  Components CPCR  39
Table 16  PCR-programs  40
Table 17  Repeat composition assembled with the GOLDEN GATE cloning strategy  42
Table 18  AvrBs3 variants  44
Table 19  Enzyme activity of AvrBs4-MutH for two independent double targets  63
Table 20  Identity of the best hits enclosing the seed region of the specific target site  75
Table 21  K_D-values of AvrBs3($\Delta$N152-C28) and AvrBs3 (B)  86
Table 22  FRET distance measurements on DNA (I) and DNA (II)  94
Table 23  Overview of the in vivo I-SceI, TALE-FokI and TALE-MutH results  105
Table 24  Overview of the nuclease activity of AvrBs3-PvuII variants  106
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>amino acid</td>
</tr>
<tr>
<td>aa</td>
<td>amino acids</td>
</tr>
<tr>
<td>AAV</td>
<td>adeno-associated virus</td>
</tr>
<tr>
<td>AD</td>
<td>activator domain</td>
</tr>
<tr>
<td>BFP</td>
<td>blue fluorescent protein</td>
</tr>
<tr>
<td>B2H</td>
<td>bacterial two-hyrid</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>CC</td>
<td>connection cilium</td>
</tr>
<tr>
<td>CRISPR</td>
<td>clustered regularly interspaced short palindromic repeats</td>
</tr>
<tr>
<td>CRD</td>
<td>central repeat domain</td>
</tr>
<tr>
<td>crRNA</td>
<td>CRISPR RNA</td>
</tr>
<tr>
<td>Cy3</td>
<td>cyanine 3</td>
</tr>
<tr>
<td>Cy5</td>
<td>cyanine 5</td>
</tr>
<tr>
<td>DMD</td>
<td>Duchenne muscular dystrophy</td>
</tr>
<tr>
<td>DNA</td>
<td>desoxyribonukleinsäure</td>
</tr>
<tr>
<td>DOL</td>
<td>degree of labeling</td>
</tr>
<tr>
<td>DSB</td>
<td>double-strand break</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>F9</td>
<td>factor IX</td>
</tr>
<tr>
<td>FGE</td>
<td>formylglycine generating enzyme</td>
</tr>
<tr>
<td>FRET</td>
<td>Förster resonance energy transfer</td>
</tr>
<tr>
<td>gal</td>
<td>galactosidase</td>
</tr>
<tr>
<td>GFP</td>
<td>green fluorescent protein</td>
</tr>
<tr>
<td>gRNA</td>
<td>guided RNA</td>
</tr>
<tr>
<td>HDR</td>
<td>homologous directed repair</td>
</tr>
</tbody>
</table>
HE  homing endonuclease
HEK 293T human embryonic kidney 293 T
HEX  hexachlorofluorescein
HGP  Human Genome Project
HR  homologous recombination
IDLV  integration-deficient lentiviral vectors
IFT  intraflagellar transport
INDELS  insertion and deletion events
IPTG  isopropyl-beta-D-thiogalactopyranosid
IS  inner segments
LB  Lysogeny broth
LCA  leber congenital amaurosis
LPLD  lipoprotein lipase deficiency
NHEJ  non-homologous end-joining
NLS  nuclear localisation signal
NTR  N-terminal region
OD  optical density
OPEN  oligomerized pool engineering
OS  outer segments
PAA  polyacrylamid
PAGE  polyacrylmid gel electrophoresis
PAM  protospacer-adjacent motif
PCR  polymerase chain reaction
PEI  polyethylenimin
PK  proteinase K
PR  photoreceptor
R  resistance
RE  restriction endonuclease
RipTAL  Ralstonia injected protein TALE
RNA  ribonucleic acid
RNF  RNA-guided FokI-dCas9 nuclease
RPE  retinal pigment epithelium
RPTG  retinitis pigmentosa GTPase regulator
RT  room temperature
RVD  repeat variable diresidue
S  host-susceptibility
SaCas9  Staphylococcus aureus Cas9
sc  single-chain
SDS  sodium dodecyl sulfate
SN2  bimolecular nucleophile substitution
SOC  super optimal broth with catabolite repression
spCas9  Streptococcus pyogenes Cas9
SSA  single-strand annealing
SSB  single-strand break
TAE  tris-acetic acid-EDTA
TBE  tris-borate-EDTA
TPE  tris-phosphate-EDTA
TALEN  transcription activator-like effector nuclease
TALE  transcription activator-like effector
TFO  triplex-forming oligonucleotid
TLR  traffic Light Reporter
tracrRNA  transactivating CRISPR RNA
UPA  upregulated by AvrBs3
UPT  upregulated by TALE
X-Gal  5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
Xcv  Xanthomonas campestris pv. vesicatoria
XLRP  X-linked form of retinitis pigmentosa
ZFN  zinc-finger nuclease
Part I

INTRODUCTION
INTRODUCTION

In 1990, the Human Genome Project (HGP) was initiated to decipher the human genome. It took until 2003 to complete sequencing of the entire human genome, which improved our understanding of genetic disease mechanisms and resulted in the development of new therapeutic strategies. One therapeutic approach was to silence dysfunctional genes with small interfering RNAs (siRNAs), resulting in the transient down-regulation of the expression specific targeted genes. However, its use is limited by the relatively low efficiency and its unfavorably safety profile [4, 114, 50]. Another approach is a therapy, which inserts the relevant gene by packing the encoding sequences into a viral-vector. A successful application of a viral-vector based therapeutic is Glybera, it treats lipoprotein lipase deficiency (LPLD) and is marketed in Europa since 2014. Glybera is injected into the skeletal muscle and introduces a healthy lipoprotein lipase gene [84, 232]. However, editing therapies carrying the risk of an immune response to the viral-vector and the non-specific insertion of the correct gene [121]. This has been shown where gene therapy was successfully employed to cure severe combined immunodeficiency-X1 (SCID X-1) via introducing a retroviral vector containing the encoding sequence of the relevant protein. However, some patients developed leukemia-like symptoms as a result of an unspecific insertion [121, 92]. Furthermore, a lethal immunological response was reported in a therapeutic trial using adenovirus vectors for the insertion of the relevant gene [216].

Homologous recombination (HR) of an exogenous DNA template with chromosomal DNA can be used to add sequences to the genome by gene correction or gene addition. However, recombination in mammalians is a rare event and therefore very inefficient. Chromosomal HR can be increased over 1000 fold by introducing double strand breaks (DSBs) in the target, thus facilitating gene targeting [61, 130, 182, 207, 181]. Various specific nucleases have been developed to introduce specific DSBs into the genome, such as meganucleases [87, 9, 201], zinc finger nucleases (ZFNs) [220, 176, 31], transcription activator-like effector nucleases (TALENs) [155, 41, 148] and the RNA-guided Cas9 nuclease [108, 140, 39, 47, 102] (Figure 1).

Site-specific DSBs trigger the cell to repair the lesion by the error prone repair mechanism non-homology end joining (NHEJ) [131] or by homology directed repair (HDR) [30, 95, 159, 60, 187]. NHEJ introduces small insertions or deletions (INDELS), leading to a disruption of the gene [131]. Thus, it can be applied to silence a pathologic gene [189, 211].
NHEJ is the more efficient repair mechanism compared to the precise homologous direct repair and is active throughout the whole cell cycle [186, 197]. The ends of the disrupted DNA are literally sticked together via NHEJ [131]. However, if a precise gene correction or gene addition is needed, a DNA template has to be introduced, which will be assembled to the DNA via HDR [61, 130, 182, 207]. However, HDR does primarily occur during the S/G2 phase of the cell cycle [43, 35] and is therefore less active and restricted to dividing cells [50].

To increase the possibility of HDR, several modifications were made [50]. For instance, suppressing the key players of NHEJ [42], but also single-strand breaks (SSBs) were introduced instead of DSBs to the DNA locus [145, 52, 222], single-stranded oligonucleotides were used instead of double-stranded substrates [36, 146], and the similarity was increased between the DNA template and the DSB site to promote formation of intermediates during the template integration [19, 154, 158].

The improvement of gene targeting is an ongoing process. Thus, the specific nucleases have to be optimized to reduce safety risks,
such as off-target activity [228, 93, 77] and a deeper understanding of
the binding mechanism could improve their specificity [142, 205, 193,
194]. Furthermore, the precise characterization of the repair mech-
anisms could optimize HDR efficiency [113]. Finally, improving the
delivery systems to the target site could be an important step towards
the effective manipulation of the human genome [121].

1.1 TOOLS FOR GENOME EDITING

The tools for genome editing are the specific nucleases, which require
a number of features to be effective and safe. They have to be highly
specific and highly efficient, but also easy to handle and to design.
The most important specific nucleases are reviewed below.

1.1.1 Meganuclease/Homing endonucleases

Homing endonucleases (HEs) are one of the first engineered nucle-
ases [61, 130, 182, 207, 181]. HEs, also known as meganuclease, target
an extremely rare large recognition site (12-40 bp) and are grouped
into five families based on sequence and structural motifs [206]. HEs
are located and encoded within protein encoding genes. HEs are self-
splicing introns (RNA) or self-splicing inteins (protein splicing). They
are considered as “selfish” nucleases as they support their own pro-
iferation. The HE induces a DSB into the foreign DNA lacking the
HE gene, triggering HDR. In turn, HDR uses the intron containing
allel as the repair matrix and the HE gene becomes duplicated [141].
I-SceI belongs to the best analyzed laclgida family, characterized
by the $\alpha_2\beta_2\alpha_2\alpha_2$ motif. Two laclgida motifs form a bipartite cata-
lytic center. Each center $\alpha$-helix contains the laclgida motif and
enclose one of the two acidic residues, which coordinate the divalent
metal ion cofactor Mg$^{2+}$ for cleavage of the phosphodiester bonds.
The four $\beta$-sheets recognize the target by forming a saddle like struc-
ture around the DNA. HEs exist as homodimer with two copies of the
laclgida motif in each monomer (e.g. I-CreI) or as monomeric
enzyme containing two laclgida motifs fused by a linker (I-SceI)
[201, 141, 171]. I-SceI recognizes a non-palidromic 18 bp sequence
and is the established golden standard due to its high specificity
combined with its high nuclease activity. Meganuclease are charac-
terized by a non-modular structure, thus engineering of their binding
domains proofed to be challenging, but can be realized by rational de-
sign and/or directed evolution [218, 106].

1.1.2 Fusion-proteins

In contrast to meganuclease, the repetitive nature of the DNA bind-
ing region of the fusion proteins ZFNs [220, 176, 31] and TALENs
enable them to address a high number of sequences. The programmable binding domains of ZFs and TALEs are typically fused to the non-specific cleavage domain of the Typ II endonuclease FokI, which functions as a dimer [117]. Improved designs have been established and are reviewed in section 1.1.4.

Another option to engineer programmable fusion proteins is to use triple-helix forming oligonucleotides (TOFs) as binding modules. TOFs build stable triple helices with the DNA through hoogsteen base pairs [63, 171]. Due to the challenging cell delivery of TOF nucleases and the slow rate of triple-helix formation, this platform of specific nucleases has no current relevance in genome editing [171].

1.1.2.1 Zinc finger nucleases

A set of zinc-fingers consist of three individual zinc-fingers and each Cys$_2$-His$_2$ finger is composed of 30 amino acids formed to a $\beta\alpha\beta$-motif. Two cysteine and two histidine residues coordinate one divalent zinc ion. Three defined amino acid residues located in the $\alpha$-helix recognize 3 bp in the major groove of the DNA [66]. A zinc-finger binding domain comprise three to four zinc-finger sets to target 9-12 bp. Several methods have been developed to construct zinc-finger proteins [3]. For example, modular assembly, which is based on a preselected library of zinc finger modules [196] or by using the oligomerized pool engineering (OPEN) method, which selects multi-finger arrays from a library and members with binding affinity are isolated with a bacterial two-hybrid (B2H) selection method [137, 136]. Unfortunately, crosstalk between adjacent zinc fingers affects the binding specificity [103] and interfere with zinc finger assembling [178]. Additionally, ZFN revealed a high off-target activity, when compared to TALEN [156].

1.1.2.2 Transcription activator-like effector (TALE)

Deciphering the code of a TAL (transcription activator-like) effector of the plant pathogenic Xanthomonas was a crucial step to improve binding domains for highly programmable specific nucleases. TAL effectors are virulence factors, which simulate eukaryotic transcription factors [25]. Gram-negative plant-pathogenic bacteria of the genus Xanthomonas inject a cocktail of effector proteins by a type III secretion system into plant cells, thus causing diseases in crop plants [115]. TAL effectors are composed of a N-terminal trans-location signal, a central repeat domain (CRD) and a C-terminal region carrying a nuclear localization signal (NLS) and a transcription activator domain [24, 26] (Figure 2 (A)). TALEs enter the nucleus and bind to a UPT (upregulated by TALE) box to activate host-susceptibility (S) genes, promoting bacterial proliferation. AvrBs3 is a well studied TALE protein from the pepper (Capsicum spec.) and tomato (Solanum lycopersicum) pathogenic strain Xanthomonas campestris pv. vesicato-
AvrBs3 interacts with the S gene UPA (upregulated by AvrBs3) inducing cellular hypertrophy of mesophyll cells [152]. Due to plant-pathogen co-evolution, disease resistance (R) genes exist in some plants. For instance, the pepper plant contains a Bs3 R gene. Here, AvrBs3 induces programmed cell death, called hypersensitive response [185].

Figure 2: 3D structure of TALE (pdb: 2YPF) [205].

(A) TALE consist of the N-terminal region (NTR) carrying the type 3 secretion signal (T3SS), the central repeat domain (e.g. 17.5 RVDs), and the C-terminus containing the nuclear localization signal (NLS) and a transcription activator domain (AD). Δ152 indicates the truncated N-terminus, while the complete C-terminus carry 278 aa after its last half repeat. (B) Crystal structure of TALE bound to DNA. TALE forms a right-handed superhelix wrapped around the major groove of the DNA. The pseudo repeats at the N-terminal region (NTR) are colored in magenta, the CRD is shown in green. (B) Magnification of one repeat. The repeat variable diresidue (RVD) at position 12 and 13 is illustrated as sticks in orange and the residues G14, K16 and Q17 are also illustrated as sticks.
Another Xcv TAL effector is AvrBs4, which is 97% identical to AvrBs3 [28]. Its corresponding R gene Bs4 is also found in some tomato plants, but can not be induced by AvrBs3 [152, 192]. Similar to the TALE protein from the plant pathogenic Xanthomonas, a TAL effector (RipTAL) from Ralstonia solanacearum with high sequence similarity to the TAL effector of Xanthomonas was identified [54], as well as a modular DNA binding protein, BurrH, from Burkholderia rhizoxinica [110, 204].

1.1.2.3 Structure and binding mechanism

The crystal structure of the binding domain of a TAL effector PthXo1 was accomplished from the rice pathogen Xanthomonas oryzae [139]. Later, AvrBs3 bound to DNA was crystallized by Stella et al. [205] (Figure 2 (B)). The central binding module is composed of highly conserved repeats, ranging from 1.5-33.5 repeats. The last repeat does always contain half of the repeat residues [24, 25]. There exist conflicting evidences regarding the necessity of the last half repeat; both a significant impact [226, 27] as well as a minor relevance [142, 248] for the binding character of TALEs have been reported. The minimum number of repeats to induce gene expression is 6.5 repeats [25] and each repeat contains 34 conserved residues. At position 12 and 13 the so called repeat variable diresidue (RVD) is located and is responsible for the “one repeat to one base” [193] specific recognition code [25, 153]. Every repeat forms two α-helices connected by a three-residue loop carrying the RVD (Figure 2 (C)). The TAL effector bound to double stranded DNA forms a right-handed superhelix wrapped around the DNA major groove [139, 205] (Figure 2 (B)). A total of 20 unique RVD sequences have been described [48], but only a few promote specific binding. RVD “HD” prefer binding to cytosine (C), RVD “NG” binds mainly to thymine (T) and RVD “NI” binds usually to adenine (A). RVD “NN” can bind to adenine (A) and to guanine (G), whereas RVD “NS” binds to adenine (A) and to cysteine (C) [25]. RVDs “NK”, “HN” or “NA” can be used to target guanine (G) [153]. The RVD “N*” missing the glycine residue at position 13, has been described to overcome the sensitivity of TALEs to methylated bases [221].

The single repeats can be assembled to an artificial functional TALE by the golden gate cloning strategy [33]. This cloning strategy enable rapid assembling of TALEs and the use of TALEs for any given target [33].

The residue at position 13 directly contacts the DNA and the residue at position 12 is responsible to stabilize the interaction between DNA and TALE [139, 57]. The residues G14, K16 and Q17 interact by non-specific polar and charge interactions with the DNA backbone [230]. Furthermore, the N-terminal region (NTR) contributes to DNA binding and consists of four continuous repeats ($N_3$, $N_2$, $N_1$, $N_0$) forming
a right handed superhelical structure similar to the CRD, but differing in its sequence [81]. Truncated versions of TALEs have been designed to reduce the overall size of the protein. It has been shown that the N-terminus missing 152 aa and the C-terminus carrying 28 aa retains its binding ability [148]. A binding sequence beginning with a thymine (Tₐ) facilitates TALE binding, but other bases can also be recognized by TALE [148, 209]. Tryptophan at position 232 in repeat N₋₁ is responsible for its Tₐ preference. By structure-guided library design, a N-terminal domain preferring guanine (Gₐ) has been constructed [124]. It has been reported that *Ralstonia solanacearum* TALE requires Gₐ and carrying an arginine at position 232 in repeat N₋₁, similar to the TALE designed by structure-guided library [124, 54].

The naturally occurring TALEs from *Xanthomonas* bind to a DNA sequence containing a 5’ T₀ [25, 153]. Interestingly, there is no naturally occurring TALE protein which matches perfectly to its target promoter. For instance, AvrBs₃ has two HD-A mismatches and one NG-C mismatch at its Bs₃ box [142]. It has been reported that TALEs from *Xanthomonas* possess single repeats with aberrant length. These repeats can be truncated or extended, when compared to the canonical repeat carrying 34 aa. The aberrant repeats allow an additional binding conformation in which the repeat loops out and therefore introduce a local flexibility to TALE-DNA binding [183].

A detailed analyses of the RVDs revealed that the overall composition of RVDs contribute to the TALE-DNA binding affinity. The order of contribution of single RVDs has been shown to be HD ≥ NG ≥ NN >>> NI > NK. Furthermore, the TALE-DNA recognition is influenced by a polarity-effect, thus the N-terminal repeats have an higher impact on the specificity and affinity of TALE compared to the C-terminal repeats [142].

Each repeat carries a conserved cysteine (C) residue at position 30, which is located in the second α-helices in a solvent-exposed outer loop [57, 139]. The cysteine residues are required to dimerize via disulfide bridges to homodimers, which occur prior to the import into the nucleus [91, 194]. However, complex formation inhibits binding to the DNA and it has therefore been suggested that the homodimers dissociate in the nucleus. Monomeric binding of TALE to its target has been shown by EMSA (electrophoretic mobility shift assays) data [194] and by the structural data [205, 139, 57] of TALE. While substitution off all cysteine residues to serine or alanine in AvrBs₃ revealed that the cysteine is required for gene activation in planta, the lack of cysteine residues slightly reduces the binding ability of TALEs under *in vitro* conditions. These results suggest that the substitution slightly affects the plasticity of AvrBs₃ [194]. X-ray crystallography revealed that TALE undergoes a conformational change upon binding to the DNA. The superhelical pitch gets reduced from 60 Å to 35 Å [57, 230]. Furthermore, conformation elasticity of TALE was
predicted by computational studies [226, 70]. Just recently, single-molecule experiments were performed with Cy3 labeled AvrBs3. The protein was labeled by hydrazine-functionalized Cy3 organic dyes at the N-terminus carrying an unnatural aldehyde and a two-state mechanism enclosing a dynamic conformation change between search and recognition state was reported [51]. The NTR binds to non-specific DNA and facilitates 1D sliding [138, 81, 51]. An extended superhelical pitch exists, which minimizes the strong interaction between CRD and the DNA backbone. This enables sliding and hopping of the TALE protein along the DNA. The single molecule results revealed an arrested state in 1D diffusion, which indicates a transition from the search state to the recognition state. The recognition state has a more compact conformation, when compared to the search state. Furthermore, they suggested that large TALEs suffer from a “slowdown” in 1D diffusion. This results in a higher frequency of transition from search to recognition state and thus results in higher off-target rates [51].

1.1.3 Clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease Cas9

The clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas)-system is an adaptive immune system in bacteria and archaea to defend the organism against invading viruses or plasmids (Figure 3) [231].

The type II CRISPR-Cas system integrates foreign sequences (protospacer) to the CRISPR cluster and transcribes these to pre-CRISPR RNAs (pre-crRNAs), which are processed to short crRNA and hybridized with transactivating CRISPR RNA (tracrRNA) and the Cas9 protein. The complex Cas9::crRNA::tracrRNA binds to its target by RNA-mediated Watson-Crick interaction and induces a DSB. To simplify the system for genome editing, a single crRNA-tracrRNA chimera was established by fusing crRNA to a truncated tracrRNA which resulted in a guide RNA (gRNA) [108]. RNA-guided Cas9 nuclease recognizes a 20 bp target which requires the protospacer adjacent motif (PAM) sequence downstream of the target for binding and cleaving activity. The commonly used Streptococcus pyogenes Cas9 (spCas9) protein primarily recognizes 5’ NGG PAM sequence [108, 107]. The Cas9 protein contains two nuclease motifs, RuvC and HNH. RuvC induces a SSB to the complementary strand and HNH induces a SSB into the non-complementary strand, resulting in a blunt end 3 bp upstream to the 5’ PAM motif [83]. The required PAM motif represents a limitation for the application of the gRNA endonuclease Cas9. However, there have been attempts to overcome the issue of the PAM motif [107, 119, 7, 72]. Partial binding of gRNA::Cas9 was demonstrated to the PAM sequence 5’ NAG [107]. The design of an engineered
Figure 3: CRISPR-Cas system

(I) Integration of a short foreign sequence (protospacer) into the CRISPR array. The protospacers are illustrated in orange, green, and red. The repeat arrays are displayed in balck. (II) Encoding tracrRNA, Cas9 protein and pre-crRNA. (III) The tracr RNA hybridizes to the repeat arrays of the pre-crRNA and (IV) are processed to short cr-RNAs. (V) Complex formation of Cas9::crRNA::tracrRNA. The complex scans the dsDNA for its target. (VI) Specific binding of the Cas9::crRNA::tracrRNA complex to the complementary protospacer sequence by Watson-Crick interaction. The PAM sequence is required to induce a DSB to the foreign DNA.

spCas9 protein with novel PAM recognition has been shown to be challenging [7], but has been recently achieved by a bacterial selection system [119]. Another attempt to improve the diversity of PAMs was the characterization of orthologues Cas9 proteins from other bacteria [119, 72]. Especially, the Staphylococcus aureus Cas9 (saCas9) revealed to be promising [119, 179]. It recognizes three novel PAM sites (NNGGGT, NNGAAT, NNGAGT) [119] and the protein saCas9 has the benefit to be ~1kb shorter compared to spCas9 [179]. The small size improves the packing features of the Cas9 protein into adeno-associated virus (AAV) vectors, a commonly applied delivery method. AAV vectors allow to pack 4.5 kb, which limits packing of the widely used spCas9 (~4.2 kb) and its gRNA into one AAV vector [179]. Additional a further CRISPR-Cas system was characterized containing a Cpf1 protein. In contrary to the Cas9 protein the Cpf1 protein does not require a transactivating CRISPR RNA, addresses a target site with a 5’ T-rich PAM sequence and induces a DSB with a 4 nt or 5 nt 5’ overhang[243].

A significant advantage of the gRNA endonuclease is the ability of re-targeting a new sequence by simply changing the sequence of the gRNA. Furthermore, the possibility to introduce multiple DSBs (mul-
tiplexing) allows new applications for the Cas9 system [47]. Nonetheless, the Cas9 nuclease does tolerate mismatches especially in the 5′ upstream region [163, 101, 76]. This limits the application of Cas9 for genome editing, especially when a high level of precision is required. To improve the specificity of RNA-guided Cas9, a nickase has been established, by introducing a point mutation in either the RuvC (D10A) or HNH (H840A) domain [108, 47]. This led to the possibility to use RNA-guided Cas9 endonuclease for inducing SSBs, but also as an inactive binding domain [174, 247].

1.1.4 Catalytic domains

1.1.4.1 Cleavage domains

Different types of nucleases have been used in fusion constructs as cleavage domains [238, 78, 22, 27]. Nucleases are phosphodiesterases hydrolyzing the phosphodiester bond, thus creating most commonly 5′-phosphates and 3′-OH groups. These proteins are essential for defending the host against invading nucleic acids [21, 175], DNA replication [111] and RNA processing [2]. Phosphodiester bond cleavage occurs due to a bimolecular nucleophile substitution (S\text{N}2), which inverts the configuration of the phosphorous. A neighboring water molecule is deprotonated to a strong nucleophile, which attacks the partially positive phosphor. Simultaneously, 3′-O′ is released. The reaction step can be catalyzed by diverse mechanisms depending on none-, one-, two-, or three-metal-ions [169, 237]. The cleavage domains analyzed in the present work belong to the two-metal-ion mechanism (FokI, PvuII and MutH) (Figure 4).

One divalent metal-ion facilitates deprotonation of the water molecule and both metal-ions stabilize the pentavalent transition intermediate. The two divalent metal-ions are coordinated by a conserved aspartate, a hallmark of the two-metal-ion mechanism [236, 169]. Homing endonucleases were also used as cleavage domains in fusion constructs. The LAGLIADAG homing endonucleases I-CreI [22] and I-Anil [27] belong to the two-metal-ion-dependent nucleases and were fused to the binding domain TALE. Furthermore, one-metal-ion dependent nucleases, like the GIY-YIG homing endonuclease I-TevI [20] and the non-specific endonucleases ColE7, NucA and EndA [22] have been shown to function as cleavage domains. The cleavage domains of the analyzed specific nucleases are reviewed in detail.
Figure 4: Cleavage domain of FokI, restriction enzyme PvuII and MutH

The essential residues of the cleavage domains are illustrated as purple sticks. One subunit is colored in blue and the second subunit in light blue. (A) Dimeric FokI (PDB: 2FOK) cleavage domains [224]. The active center is build by the residues D450, D467 and K468. (B) PvuII (PDB:1Pvl) with its catalytic residues D58, E68 and K70 in each subunit [99]. Naturally, the PvuII restriction enzyme function as homodimer. To design a heterodimeric PvuII, the substitution D30K or K38D was introduced to one subunit, respectively. D30 and K38 are displayed as orange sticks. In red the linker (Gly-Ser-Gly-Gly) used to fuse the identical subunits of PvuII and to create a scPvuII variant is shown. (C) Monomeric nickase MutH interacting with its target DNA. (PDB: 2AOQ) Catalytic residues are D70, E77 and K79 [125]. Double stranded DNA is demonstrated in gray.

1.1.4.2 FokI

The endonuclease FokI, isolated from Flavobacterium okeanokoites, belongs to the type IIS restriction endonucleases.

The enzyme recognizes a 5 bp asymmetric nucleotide sequence, cleaving 9 (top-strand) and 13 (bottom-strand) nt downstream of the site (5’-GGATG(N)9 NNNWΔ13) [208].

Two domains shape the FokI restriction endonuclease, a N-terminal binding domain (41 kDa) and a C-terminal cleavage domain (25 kDa)
The two domains are fused by a flexible linker and the cleavage domain is kept in an compact conformation until binding to the substrate [167]. Binding to the DNA target induces a conformational change, which recruits the catalytic cleavage domain 13 nt downstream of the site to the bottom-strand. Next, a second monomeric FokI enzyme dimerizes by protein-protein interaction and is responsible to cut the top-strand 9 nt away from the recognition site [167, 188, 23].

The second monomeric FokI endonuclease can be bound to a proximate second recognition site or is available from the free solution [93]. However, the dimerization between bound and free FokI proteins is unstable, resulting in a low cleavage rate for one-target site DNA [23]. Unless free FokI enzyme concentration is >>100 nM, since the equilibrium dissociation constant of the second FokI monomer is ≈100 nM [32]. Consequently, the catalytic domain of FokI is an unspecific cleavage domain, which requires dimerization to induce DSBs [224, 23] (Figure 4 (A)).

The catalytic domain of FokI was fused to the C-terminus of a zinc-finger domain by the Chandrasegaran group, creating the first reported ZFN [117]. Thereafter, improved ZF-FokI designs were created. One approach to increase the specificity was to design heterodimeric catalytic domains by introducing defined mutations [147, 212, 90]. Furthermore, sharkey, a FokI catalytic domain with increased activity was designed and is reported to be compatible with the heterodimeric mutations [90]. Another approach to optimize the construct was the design of the single-chain (sc) ZN-FokI specific nuclease. Two copies of the FokI catalytic domain were fused with a flexible (Gly-Gly-Gly-Gly-Ser)$_n$ polypeptide linker to each other [149, 213]. When TALE, as a new binding domain was discovered, the cleavage domain FokI was applied to the new design [33, 129, 148]. Likewise to the scZN-FokI design, a scTALE-FokI [210] construct with a 95 aa polypeptide linker was constructed. Recently, the unspecific catalytic domain of FokI was fused to the amino-terminus of a catalytic inactive Cas9 protein to improve the off-target character of the CRISPR/-Cas9 -system [89, 219].

To enable FokI dimerization, two adjacent sites are placed 15-39 bp apart from each other with PAM sites facing outside. This new dimeric RNA-guided FokI-dCas9 nuclease (RNF) has been shown to improve the sensitivity to mismatches [8, 235].

1.1.4.3 PvuII

The restriction enzyme PvuII (Figure 4 (B)), isolated from Proteus vulgaris, belongs to the type IIIP restriction enzymes. PvuII addresses a palidromic 6 bp sequence and induces a DSB through dimerization, resulting in blunt ends (5'-CAG\footnote{5'}CTG-3'). Each PvuII subunit contains a PD...D/ExK motif, characterized by the crucial aa of the
active center (D58, E68 and K70) [37, 170, 169]. The advantage of PvuII over the unspecific catalytic domain FokI is its additional site-specific sequence specificity [239]. Similar to the catalytic domain FokI, the PvuII enzyme was optimized for the use as cleavage domain. Thus, a scPvuII version has been designed by fusion of two identical PvuII subunits with a short aa linker (Gly-Ser-Gyl-Gly) [202, 144]. The first alternative architecture of programmable nucleases with the PvuII cleavage domain was the triple-helix-formation oligonucleotide (TFO)-scPvuII fusion construct [65]. However, this design has to be pre-incubated in the absence of Mg^{2+} metal-ions and activated with Mg^{2+} metal-ions to cleave the target site specifically. This required activation process complicates the application of TFO-scPvuII in vivo [239]. Furthermore, PvuII was fused to inactivated homing endonuclease I-SceI* [71], to a zinc-finger array [190] and to the TALE binding domain [239, 238]. The new site-specific architecture of programmable nucleases with PvuII proofed to be superior with the fusion constructs with FokI compared to zinc-finger-PvuII and zinc-finger-FokI [190]. PvuII connected to the truncated binding domain TALE (ΔN152-C28) was analyzed as homodimer, as monomeric single-chain construct and as heterodimer. Here, PvuII contained the T46G mutation, which is known to reduce “star-activity” and has been termed as high fidelity (hf) PvuII [250]. “Star-sites” are similar, but not identical to the recognition site. To reduce the toxicity of TALE-PvuII the cleavage domain PvuII was further optimized. Ten new PvuII variants, each with a novel single mutation, were generated by directed evolution. Five of these new variants retained the cleavage activity of the fusion construct. Among these five variants, the TALE-PvuII(T46G/G135W) variant revealed the best balance between activity and toxicity. Toxicity was analyzed in different Escherichia coli (E. coli) strains (JM109 and XL10Gold), as well as in an eucaryotic cell line (HEK293) [239, 238]. In addition, heterodimer PvuII subunits were designed to increase the specificity. Therefore, PvuII subunits were created containing a D30K mutation or a K38D mutation. These aa interact with each other during the dimerization of the PvuII subunits. The mutations are located centrally, but are not placed in the active center. Due to the different charges of PvuII(D30K) and PvuII(K38D), exclusively heterodimeric PvuII subunits can dimerize and induce a DSB [238].

### 1.1.4.4 Programmable nickases

To further reduce off-target activity, nucleases were modified by the utilization of nickases to induce SSBs rather than DSBs, thus favoring the occurrence of HDR over the error prone repair mechanism NHEJ [145, 52, 222].

Programmable nickases have been created by (i) inactivation of one of the binary catalytic cleavage centers (e.g. N450A FokI [116, 188,
introduction

177, 227], K223I-SceI [157, 113] and D10A CAS9 Mali et al. [140], Ran et al. [180], Cho et al. [40]), or (ii) by fusion of a naturally occurring nickase to its binding domain, e.g. TALE-MutH [78]. Nickases have been designed to induce SSBs in the sense or anti-sense strand. This feature can be exploited to increase HDR efficiency, as it is crucial when the transcribed or the non-transcribed strand are nicked [53]. Furthermore, specific nickases have been used for double nicking (DN), thus inducing SSBs on both DNA strands, which resulted in an enhanced genome editing specificity [191, 199, 180, 40]. It has been shown by Ran et al. [180] that the efficiency of double nicking depends on the offset between the two target sites and on the created overhang. HDR efficiency triggered by paired Cas9 nickases has been reported to be comparable with the HDR efficiency induced by wt Cas9 endonuclease creating DSBs. Generating a 5’overhang does improve NHEJ and HDR efficiency [40, 180, 140]. While the offset can be up to 100 bp, it is optimal up to 20 bp. Finally, at least one nick should occur in the homologous part to optimize the efficiency and no improvement of HDR efficiency could be detected by generating a 3’overhang [180, 140].

1.1.4.5 MutH

MutH (Figure 4 (C)) is a 28 kDa monomeric site and strand specific latent endonuclease which requires activation by MutS and MutL associated with a mismatched base pair [125, 12]. It induces a SSB in un- and hemi-methylated strands 5’ to its recognition site \(\nabla\)GATC. The structure of MutH is surprisingly similar to the PvuII endonuclease. Furthermore, the active center of MutH contains the core aa Asp70, Glu77 and Lys79 (the DEK triad), which is also found in the PD...D/ExK motif of PvuII [12]. The monomeric endonuclease belongs to the mismatch repair system in system in E. coli and some other bacteria. MutS recognizes the mismatch and recruits MutL [94]. This ternary complex activates MutH, which induces a SSB at the 5’-side of a hemimethylated GATC site. The mismatch can be >1000 bp upstream or downstream of the GATC site [122]. Next, single-stranded DNA is generated and digested by an exonuclease across the original mismatch [151]. Without recruitment, the nickase MutH is barely active or inactive under physiological conditions [229, 64]. This feature was exploited by Gabsalilow et al. [78], as they fused the monomeric site and strand specific endonuclease to the binding domain TALE.

TALE-MutH is a strand- and site-specific programmable nickase, which has been characterized in vitro [78]. Dependent on the spacer length between the TALE binding site and the GATC site, the SSB is induced in the top- or bottom-strand. Targets containing a spacer of 3 bp are nicked on the bottom-strand, whereas the optimal spacer length for the top-strand cleavage is 6 bp [78]. This feature allows to
specifically target either the transcribed or the non-transcribed strand by TALE-MutH, which has been shown to be crucial in respect of HDR efficiency and therefore successful genome editing [53].

1.1.4.6 Diversity of the catalytic domains

Beyond the function as specific nuclease the described “tools for genome editing” were also modified for other applications. For instance, TALE proteins were fused to activator (e.g. vp16) [245, 152, 148] or repressor domains (e.g. KRAB) [82] to activate gene expression or to down-regulate or even arrest gene expression. Other designed fusion constructs were realized with DNA-methyltransferases [18], with demethylases [134], or with histon-modifying enzymes [143, 38]. To catalyze DNA integration, excision, and inversion recombinases [80, 85], or transposases [240] were fused to the binding domain TALE [79]. Additionally, TALEs were fused to fluorescent proteins to visualize target sequences [150]. Optical control of transcription was achieved by an optogenetic two-hybrid system, employing the light-inducible heterodimerizing proteins CRY2 and CIB1. This was realized by fusing CIB1 to the transactivation domain VP64 and CRY2 to TALE [120] or to the inactive Cas9 protein [172]. In the presence of blue light, a conformation change of CRY2 occurs and CIB1 forms a heterodimer with CRY2. Furthermore, inactive Cas9 protein was applied to repress transcription by blocking the RNA polymerase [174]. These examples underline the high relevance of the discussed proteins and their diverse applications.

1.2 Therapeutic applications

A major limitation of the application of specific nucleases for genome editing is their specificity. Studies compared the specificity of the tools and revealed that the binding proteins tolerate mismatches depending on location and composition [142, 40]. For instance, TALE proteins are more sensitive to mismatches at the 5’-end of the target [142], while Cas9:gRNA tolerates mismatches at the 5’-end better than at the 3’-end [108, 47, 107]. TALENs have been shown to be the best choice when high accuracy is required (e.g. for gene correction) [155, 156, 228]. However, by choosing unique target sites [40], the possibility of double nicking with Cas9 proteins [40] and the fusion construct Cas9:FokI [219] increased the specificity and the field of application for Cas9 proteins. However, other factors affect the safe and efficient application of specific nucleases.

1.2.1 Delivery systems

The gene editing technologies have to be delivered into the target cell for their successful application. This can be realized by *ex vivo* or *in*
**vivo** editing, using viral or non-viral vector-mediated delivery or by employing physical delivery methods. The viral or non-viral vectors carry the expression cassette for the specific nucleases and in some case also the DNA template. The proteins can also be introduced to the cell as purified proteins. For **ex vivo** editing therapy, the target cell population is removed from the organ, modified and then re-engrafted into the body. This enables the usage of a wide range of delivery methods and a control of the dosage of the delivered molecules. However, **ex vivo** editing is limited to certain tissues, as the cell population needs to survive the manipulation outside the body and has to be able to return back to the tissue after transplantation. In **in vivo** therapies, the gene editing technologies are introduced directly into the diseased cells. This allows to target tissues not amenable for **ex vivo** editing therapy and the treatment of diseases affecting multiple organ systems [50]. However, major drawbacks of **in vivo** editing therapies are the possibility of an immune response to the viral-vector and the non-specific insertion of the corrected gene [121].

1.2.1.1 **viral vectors:**

Integration-deficient lentiviral vectors (IDLVs) and adeno-associated virus (AAV) are tools to deliver nucleases **ex vivo** and **in vivo**. They are characterized with high transduction efficacy and long-term gene expression. IDLVs provide the ability to pack the specific nuclease and the donor DNA for HDR, but showed to integrate randomly into the host genome. AAV vectors are non-pathogenic and facilitate delivery into dividing and non-dividing cells. Though, AAV vectors are limited by their carrying capacity (~4.5 kb). ZFNs are relatively small (~1kb) and thus suitable for AAV vectors. However, TALENs can only be packed as monomers with a small promoter sequence into the AAV vectors. In the case of the spCas9 protein, a small orthogonal saCas9 was developed, which can be packed with gRNA to the AAV vector. Additionally, the repetitive nature of TALE-encoding genes are an interference, as rearrangements and truncations occur using both viral vectors [132, 121].

1.2.1.2 **non-viral vectors:**

Non-viral vectors possess a high carrying capacity, less toxicity and are easily produced. But they revealed to be less efficient and also toxic at high concentrations. Cationic liposomes and cationic polymers such as polyethylenimine (PEI) are commonly used non-viral vectors [3]. The positive charge interact with the anionic nature of the cell membrane and triggers cellular uptake [121]. Furthermore, direct delivery of purified proteins was realized. Due to the positive charge of the zinc fingers, they are cell-penetrating proteins. Modified TALENs with cell-penetrating peptides do also enable cellular
uptake. A similar approach was used to promote cellular uptake of the Cas9 protein and its gRNA [132].

1.2.1.3 **Physical methods:**

The most commonly used physical methods for delivery of specific nucleases are electroporation and microinjection. Electroporation generates microscopic pores due to electrical pulses. By microinjection the editing technology can be introduced directly to the cell, using a needle (0.5 μm -20 μm in diameter) in combination with a specialized microscope [3].

1.2.2 **Impact of epigenetics on enzyme activation**

The accessibility of the target sequence in the chromatin structure is another aspect to consider. Relaxing the chromatin structure by the inhibition of epigenetic modulatory activity of the histone deacetylases improved the targeting frequency of specific nucleases [29]. Silencing a protein encoding gene which is required for chromatin remodeling has been shown to improve the efficiency of specific nucleases [56]. Additionally, blocking DNA-methyltransferase resulted in an increased TALENs activity [29]. Consequently, an optimized TALE RVD was designed to overcome the sensitivity to cysteine methylation [221]. Due to the epigenetic mechanisms of the cell cycle, the chromatin structure is looser in the S phase, when DNA replication occurs [13]. Thus, optimized cleavage takes place in the S phase of the cell cycle [164].

1.2.3 **Therapeutic potential**

ZFNs are the most commonly used therapeutic tool, as they are the first developed specific nucleases [117]. ZNFs addressing the CCR5 gene to treat Human immunodeficiency virus (HIV-1) resistance have already completed an early phase of human clinical trails. The CCR5 gene is disrupted by NHEJ, which prevents expression of the CCR5 receptor. This receptor is needed for viral uptake to the CD4+ T cell, which serves the virus as host cell [97, 166, 214]. TALENs and Cas9 proteins were employed to treat HIV-1 in different cell lines [241, 68]. Another clinical relevant example is hemophilia B, a genetic disorder causing a deficiency of the coagulation factor IX (F9). Introducing a donor DNA with the expression sequence of factor IX by HDR increased the concentration of F9 and improved clotting time in transgenic hemophilic mice [126]. Treatment of the X-linked hereditary disease, Duchenne muscular dystrophy (DMD) was realized with ZFNs [161], TALENs [161] and the CRISPR/Cas9 system [133]. The monogenic disease is caused by a mutation in the dystrophin gene, which is required for muscular function. Treatment was realized with ZFNs,
an ex vivo protocol, and manipulated undifferentiated cells were transplanted back to the limb of mice [161]. Furthermore, the newly established CRISPR/Cas9 system was used to treat DMD in mice [133].

Recently, a 11 month girl with B cell leukemia was successfully treated with universal chimeric antigen receptor (UCAR)19 T-cells. The UCAR19 T-cells were engineered by the TALEN-based editing technology of Cellectis to suppress the alloreactivity of the T-cells and acquire resistance to the monoclonal antibody Alemtuzumab, which is a drug to treat B cell leukemia [173]. These examples demonstrate the therapeutic potential of specific nucleases and the diversity of their genetic targets [121].

### 1.2.3.3 Gene therapy without specific nucleases

#### 1.2.3.3.1 Ocular gene therapy

Treatment of photoreceptor (PR) degeneration in the eye by gene addition has been lately achieved [105, 11, 44]. The eye represents an excellent target for gene therapy [14]. Three independent clinical trials revealed the efficiency of gene therapy in patients with RPE65-associated leber congenital amaurosis (LCA) (NCT00481546, NCT00516477, NCT00643747, clinicaltrials.gov). RPE65-associated LCA is caused by RPE65 deficiency, an isomerase in the retinal pigment epithelium (RPE) [44, 234]. RPE is a part of the blood-retinal barrier and in metabolic relationship with the adjacent PR layer of the neural retina.

The neural retina consists of the outer segment (OS) and the inner segment (IS), which are linked by the connection cilium (CC). The OS is light sensible and contains rods and cones, which are packed with the G-protein-couple receptor opsin [233, 100]. The IS is responsible for the protein production and the CC transports the proteins from the IS to the OS by intraflagellar transport (IFT) [100]. The light-sensitive chromophore 11-cis retinal is covalently bound to the G-protein-couple receptor resulting rhodopsin (RHO) (Figure 5).

Activation of rhodopsin occur by light due to isomerization of 11-cis retinal inducing a signal cascade [233]. Patients with RPE65 deficiency have degeneration of RPE65 and PR, due to the missing isomerase which results in the accumulation of all-trans retinal [233]. By introducing a recombinant AAV carrying the complete RPE65 gene, the retinal sensibility could be improved [45, 44, 11, 105].
1.2.3.3 X-linked form of retinitis pigmentosa (XLRP)

The successful application of gene therapy in RPE65-associated leber congenital amaurosis resulted in attempts to treat other PR degenerative diseases with gene therapy, such as the X-linked form of retinitis pigmentosa (XLRP). XLRP represents 22% of all forms of retinitis pigmentosa and 80% of all XLRP cases are caused by a mutation in the retinitis pigmentosa GTPase regulator (RPGR)-gene [100, 223, 198, 165]. The RPGR is concentrated in the CC, as well as in the OS of vertebrate PRs and plays an essential role for CC trafficking. In fact, RPGR can be found in many cells of the body in distinct splicing variants of the gene. The RPGR\textsubscript{ORF15} splicing variant contains a terminal exon with a repetitive glutamic acids rich domain, called open reading frame 15 (ORF15) and is found in large quantities in the PRs [100]. The ORF15 consists of an imperfect repeat motif, which carries 80% of sequence alteration, causing a dysfunctional RPGR [223, 100]. Opsin misslocation is often a hallmark for RPGR dysfunction [217, 234, 223].

Lately, AAV2/5 vectors carrying a full-length human RPGR\textsubscript{ORF15} cDNA were subretinally injected to four dogs which carried two distinct RPGR mutations [15]. Furthermore, the full-length RPGR\textsubscript{ORF15} packed into AAV 8 or 9 capsin were subretinally injected into Rpgr-KO mice and could rescue the retinal structure and function [234].

![Figure 5: The outer segment of the photoreceptor cells.](image)

The outer segments (OS) are adjacent to the retinal pigment epithelium (RPE). The light-sensitive chromophore 11-cis retinal is covalent bound to opsin in the OS. Rhodopsin is activated by light, which trigger isomerization of 11-cis retinal inducing a signal cascade. The resulting all-trans retinal gets released from the rhodopsin and then it is transported to the RPE. In the RPE it is modified to all-trans retinol, while it result in 11-cis retinal by two following enzymatic catalyzed steps requiring RPE65 and a dehydrogenase.
1.3 AIM

The aim of the present study was to characterize specific nucleases with a focus on the site- and strand-specific nucleases TALE-MutH as well as the binding domain AvrBs3.

We claim to analyze the mechanism of nicking and double nicking by TALE-MutH in vitro as well as its enzymatic activity under in vivo conditions using a Single-Strand-Annealing (SSA) assay in *Saccharomyces cerevisiae*. In a comprehensive study the characteristics of the monomeric nickase TALE-MutH, I-SceI and the widely used TALE-FokI will be analyzed. Additionally, the characteristics of TALE-PvuII variants and the gRNA nuclease under in vivo condition employing the SSA assay in *S. cerevisiae* should be defined and compared to I-SceI. The most efficient specific nucleases will be challenged by targeting a relevant therapeutic gene, such as the RPGR gene.

Furthermore, the thesis focused on the binding domain AvrBs3 and aims to define the binding mechanism of TALE. There are several studies characterizing the binding mechanism of TALE, while the exact binding mechanism still remains to be elucidated. A hindrance to investigate the binding mechanism are the lacking or restricted methods. Here, we claim to design an AvrBs3 which enable fluorescent labeling at specific positions, while retaining the binding characteristics. In 2015 AvrBs3 was labeled with Cy3 via an unnatural aldehyde and for the first time the in vitro binding character of AvrBs3 was described with single molecule studies [51]. In the present study, we want to confirm the findings of Cuculis et al. [51] with a focus on the conformation change between a non-specific TALE::DNA complex and a specific TALE::DNA complex. This should be achieved by Förster resonance energy transfer (FRET) experiments, which depend on the distances between donor fluorophore and acceptor fluorophore. Two single labeled AvrBs3 variants should be designed, which allows labeling of AvrBs3 specific and close to the N-terminus region or close to the C-terminus region. We were aspired to design an AvrBs3 labeled at two specific positions enabling to follow conformational changes within the protein.
Part II

MATERIAL AND METHODS
MATERIAL AND METHODS

2.1 MATERIALS

2.1.1 Chemical

Chemicals were applied from MERK, ROTH, SIGMA-ALDRICH, APPLICHEM or QIAGEN and were diluted in molecular biology grade water.

2.1.2 Commercial DNA purification kits

Kits used for DNA purification are listed in table 1.

<table>
<thead>
<tr>
<th>Table 1: Commercial purification kits for DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial purification kit</td>
</tr>
<tr>
<td>Wizard R SV Gel and PCR Clean – up System</td>
</tr>
<tr>
<td>PureYieldTM Plasmid Miniprep System</td>
</tr>
<tr>
<td>PureYieldTM Plasmid Midiprep System</td>
</tr>
</tbody>
</table>

2.1.3 Columns/matrices

By the principle of affinity chromatography the recombinant proteins were purified above Strep-, His- beads and/or Heparin columns. Labeled recombinant proteins were purified by the principle of size-exclusion chromatography with Zeba Desalting spin columns. Table 2 gives the employed column and matrices.

<table>
<thead>
<tr>
<th>Table 2: Columns and matrices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Column / matrix</td>
</tr>
<tr>
<td>Ni-NTA agarose</td>
</tr>
<tr>
<td>Strep-Tactin R Sepharose R</td>
</tr>
<tr>
<td>HiTrap Heparin HP</td>
</tr>
<tr>
<td>Zeba Desalting spin columns (40K)</td>
</tr>
</tbody>
</table>
2.1.4 *Dialyse tube*

For dialyse the dialysis tube Visking type 8/32 (Roth) was used. The semi-permeable membrane has an exclusion size of 14 kDa. The tubes were boiled in ultra-pure water and stored in 20% ethanol at 4°C. Before use the tubes were washed with ultra-pure water and dialyse buffer.

2.1.5 *Buffers and media*

In this section, all employed buffers and media are listed. Buffers and media were prepared with ultra-pure water from Mili-Q A10 ultra-pure water system (Milipore).

Buffer used for activity assays, for labeling reaction, as well as for anisotropy and FRET experiments are given in table 3.

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>low salt activity buffer</td>
<td>20 mM tris - acetate</td>
</tr>
<tr>
<td>50 mM potassium acetate</td>
<td>pH 7.9</td>
</tr>
<tr>
<td></td>
<td>50 mM potassium acetate</td>
</tr>
<tr>
<td></td>
<td>10 mM magnesium acetate</td>
</tr>
<tr>
<td></td>
<td>1 mM DTT</td>
</tr>
<tr>
<td>labeling buffer</td>
<td>10 mM HEPES-KOH</td>
</tr>
<tr>
<td></td>
<td>1 mM EDTA</td>
</tr>
<tr>
<td></td>
<td>200 mM potassium chloride</td>
</tr>
<tr>
<td></td>
<td>10% Glycerin</td>
</tr>
<tr>
<td>physiological activity buffer</td>
<td>20 mM tris - acetate</td>
</tr>
<tr>
<td></td>
<td>pH 7.5</td>
</tr>
<tr>
<td></td>
<td>120 mM potassium acetate</td>
</tr>
<tr>
<td></td>
<td>1 mM magnesium acetate</td>
</tr>
<tr>
<td></td>
<td>0.1 M DTT</td>
</tr>
<tr>
<td>1x NEB 3.1 buffer</td>
<td>50 mM Tris-HCl</td>
</tr>
<tr>
<td></td>
<td>pH 7.9</td>
</tr>
<tr>
<td></td>
<td>100 mM NaCl</td>
</tr>
<tr>
<td></td>
<td>10 mM MgCl2</td>
</tr>
<tr>
<td></td>
<td>100 ug/ml BSA</td>
</tr>
</tbody>
</table>
To identify, quantify, and purify nucleic acid fragments, DNA electrophoresis was used. The employed buffers are recorded in table 4.

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 x Tris-phosphate-EDTA (TPE) buffer</td>
<td>0.9 M Tris, pH 8.2, phosphoric acid, 20 mM EDTA</td>
</tr>
<tr>
<td>10 x Tris-borate-EDTA (TBE) buffer</td>
<td>1 M Tris, pH 8.3, 1 M boric acid, 25 mM EDTA</td>
</tr>
<tr>
<td>50 x Tris-acetic acid-EDTA (TAE) buffer</td>
<td>2 M Tris, pH 8.0, 1 M sodium acetate, 50 mM EDTA</td>
</tr>
<tr>
<td>10 x Laemmli-SDS-electrophoresis-buffer</td>
<td>250 mM Tris, pH 8.3, 1.9 M glycine, SDS 1% (w/v)</td>
</tr>
<tr>
<td>5 x AAP agarose loading buffer</td>
<td>250 mM EDTA, pH 8.0, sucrose 25% (w/v), SDS 1.2% (w/v), bromphenol blue 0.1% (w/v)</td>
</tr>
<tr>
<td>6 x SDS loading buffer</td>
<td>420 mM Tris-HCl, pH 6.8, 0.1% SDS, 2-mercaptoethanol 7.2 (v/v), glycerol 36% (v/v), bromphenol blue 0.06% (w/v)</td>
</tr>
<tr>
<td>Coomassie staining solution</td>
<td>coomassie brilliant blue G250, 0.1% (w/v), phosphoric acid 2% (v/v), aluminum sulfate 5% (w/v), ethanol 10% (v/v)</td>
</tr>
</tbody>
</table>
Table 5 provides buffers used for working with *Saccharomyces cerevisiae*.

<table>
<thead>
<tr>
<th>BUFFER</th>
<th>COMPONENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>transformation mix</td>
<td>0.1 M lithium acetate</td>
</tr>
<tr>
<td></td>
<td>10 mM tris-HCl</td>
</tr>
<tr>
<td></td>
<td>1 mM EDTA</td>
</tr>
<tr>
<td></td>
<td>polyethylene glycol (PEG 3350) 40% (v/v)</td>
</tr>
<tr>
<td>LTE buffer</td>
<td>0.1 M lithium acetate</td>
</tr>
<tr>
<td></td>
<td>10 mM tris-HCL</td>
</tr>
<tr>
<td></td>
<td>1 mM EDTA</td>
</tr>
<tr>
<td>10 x BU salt</td>
<td>250 mM NaH$_2$PO$_4$</td>
</tr>
<tr>
<td></td>
<td>pH 7</td>
</tr>
<tr>
<td></td>
<td>195 mM Na$_2$HPO$_4$ 7 H$_2$O</td>
</tr>
<tr>
<td>10 x dropout solution</td>
<td>200 mg/L L-Adenine hemisulfate salt</td>
</tr>
<tr>
<td>missing appropriate amino acid</td>
<td>200 mg/L L-Arginine HCl</td>
</tr>
<tr>
<td></td>
<td>200 mg/L L-Histidine HCl monohydrate</td>
</tr>
<tr>
<td></td>
<td>300 mg/L L-Isoleucine</td>
</tr>
<tr>
<td></td>
<td>1000 mg/L L-Leucine</td>
</tr>
<tr>
<td></td>
<td>300 mg/L L-Lysine HCl</td>
</tr>
<tr>
<td></td>
<td>200 mg/L L-Methionine</td>
</tr>
<tr>
<td></td>
<td>500 mg/L L-Phenylalanine</td>
</tr>
<tr>
<td></td>
<td>2000 mg/L L-Threonine</td>
</tr>
<tr>
<td></td>
<td>200 mg/L L-Tryptophan</td>
</tr>
<tr>
<td></td>
<td>300 mg/L L-Tyrosine</td>
</tr>
<tr>
<td></td>
<td>200 mg/L L-Uracil</td>
</tr>
<tr>
<td></td>
<td>1500 mg/LL-Valine</td>
</tr>
</tbody>
</table>
Media used for \textit{E. coli} and \textit{S. cerevisiae} are given in table 6.

<table>
<thead>
<tr>
<th>MEDIA</th>
<th>COMPOUNDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysogeny broth (LB) media</td>
<td>trypton 1% (w/v)</td>
</tr>
<tr>
<td></td>
<td>yeat extract 0.5% (w/v)</td>
</tr>
<tr>
<td></td>
<td>NaCl 0.5% (w/v)</td>
</tr>
<tr>
<td></td>
<td>pH 7.5</td>
</tr>
<tr>
<td>SOC media</td>
<td>bacto tryptone 2% (w/v)</td>
</tr>
<tr>
<td></td>
<td>yeast extract 0.5% (w/v)</td>
</tr>
<tr>
<td></td>
<td>10 mM NaCl</td>
</tr>
<tr>
<td></td>
<td>0.5 mM potassium chloride</td>
</tr>
<tr>
<td></td>
<td>10 mM MgCl$_2$</td>
</tr>
<tr>
<td></td>
<td>10 mM MgSO$_4$</td>
</tr>
<tr>
<td></td>
<td>20 mM glucose</td>
</tr>
<tr>
<td></td>
<td>pH 6.5</td>
</tr>
<tr>
<td>YPAD media</td>
<td>yeast extract 1% (w/v)</td>
</tr>
<tr>
<td></td>
<td>bacto pentone 2% (w/v)</td>
</tr>
<tr>
<td></td>
<td>dextrose 2% (w/v)</td>
</tr>
<tr>
<td></td>
<td>bacto agar 2% (w/v)</td>
</tr>
<tr>
<td></td>
<td>adenine hemisulfate salt 7.5 x 10$^{-3}$% (w/v)</td>
</tr>
</tbody>
</table>
To grow *E. coli* colonies LB-plates containing antibiotics were used. For *S. cerevisiae* drop out plates were employed (Table 7).

<table>
<thead>
<tr>
<th>plate</th>
<th>compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>LB-plate</td>
<td>LB-media</td>
</tr>
<tr>
<td></td>
<td>agar 0.015 (w/v)</td>
</tr>
<tr>
<td>drop out plate (-HIS/-TRP)</td>
<td>yeast nitrogen base without aa 0.67% (w/v)</td>
</tr>
<tr>
<td></td>
<td>agar 0.0% (w/v)</td>
</tr>
<tr>
<td></td>
<td>1 x dropout solution (-HIS/-TRP)</td>
</tr>
<tr>
<td></td>
<td>dextrose 2% (w/v)</td>
</tr>
<tr>
<td></td>
<td>0.08 g/ml X-GAL</td>
</tr>
<tr>
<td></td>
<td>1 x BU salt buffer</td>
</tr>
<tr>
<td>drop out plate (-His)</td>
<td>yeast nitrogen base without aa 0.67% (w/v)</td>
</tr>
<tr>
<td></td>
<td>agar 0.03% (w/v)</td>
</tr>
<tr>
<td></td>
<td>1 x dropout solution (-HIS)</td>
</tr>
<tr>
<td></td>
<td>dextrose 2% (w/v)</td>
</tr>
<tr>
<td></td>
<td>1 x BU salt buffer</td>
</tr>
</tbody>
</table>
2.1.6 Fluorophores

Several fluorophores were used to either label target DNA or different variants of AvrBs3 proteins and are listed in table 8.

Table 8: Physical properties of fluorophores used for labeling DNA and proteins. Maleimide-fluorophores were utilized for protein labeling and reacted with the cysteine residues of the AvrBs3 variants. $A_{F,max.}$ and $E_{F,max.}$: Absorbance and fluorescence emission maxima in nm; * correction factor of fluorophores at 280 nm and 260 nm.

<table>
<thead>
<tr>
<th>Physical properties</th>
<th>Cy3</th>
<th>Cy5</th>
<th>HEX</th>
<th>Alexa594</th>
<th>Alexa488</th>
<th>Alexa647</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{F,max.}$ [nm]</td>
<td>550</td>
<td>649</td>
<td>535</td>
<td>590</td>
<td>495</td>
<td>650</td>
</tr>
<tr>
<td>$E_{F,max.}$ [nm]</td>
<td>570</td>
<td>670</td>
<td>560</td>
<td>617</td>
<td>519</td>
<td>665</td>
</tr>
<tr>
<td>$\varepsilon_{max}$ [cm$^{-1}$M$^{-1}$]</td>
<td>150,000</td>
<td>250,000</td>
<td>96,000</td>
<td>73,000</td>
<td>71,000</td>
<td>239,000</td>
</tr>
<tr>
<td>cf$_{280}$ *</td>
<td>-</td>
<td>0.05</td>
<td>-</td>
<td>0.56</td>
<td>0.11</td>
<td>0.030</td>
</tr>
<tr>
<td>cf$_{260}$ *</td>
<td>0.08</td>
<td>-</td>
<td>0.33</td>
<td>0.43</td>
<td>0.30</td>
<td>0.00</td>
</tr>
</tbody>
</table>

In figure 6 the chemical structure of the cyanine fluorophores Cy3 (I) and Cy5 (II) are given, as well as the chemical structure of the 6-carboxy-2,4,4,5,7,7-hexachlorofluorescein succinimidyl ester HEX (II) fluorophore. Cy3 and HEX fluorophores were labeled to the DNA and Cy5 maleimide was used to label the single cysteine AvrBs3 variants.

![Chemical structures of fluorophores](image)

Figure 6: Chemical structures of fluorophores

Chemical structure of the fluorophores Cy3 (I), Cy5 (II) and HEX (III).
The chemical structures of the fluorophores used for labeling the single cysteine variants, as well as the double cysteine AvrBs3 variants are given in figure 7. All fluorophores used to label the protein contained a maleimide functional group.

![Chemical structures of fluorophores](image)

**Figure 7**: Chemical structures of fluorophores

Chemical structure of Alexa 488 (I), Alexa 594(II), and Alexa 647(III).

### 2.1.7 Plasmids

Plasmids were used to construct the binding domains TALE or as expression and substrate vector.

#### 2.1.7.1 Binding domains TALE

Binding domains TALE were cloned by the Golden Gate TALEN and TAL effector kit 2.0. This plasmid Kit was a gift from Daniel Voytas and Adam Bogdanove (Addgene kit # 100000024) [33]. It contains all required plasmid to construct a TALE binding domain.

#### 2.1.7.2 Expression vectors

**pQE – 30 vector** The pQE-30 vectors can be used for protein expression in all ampicillin - sensitive E. coli strains containing pREP4 repressor or lacZ gene on the F-factor episome. The vector harbours a Ts – promoter, which is under control of the lac-repressor. The N-terminal His-tag was sub cloned to the C-terminus coding sequence and a Strep-tag was added to the N-terminal coding sequence.
PQE-30_strep-avrBs3(N152-C28) contains the truncated version of AvrBs3 (ΔN152-C28).

PQE-30_strep-avrBs4 (28)-mutH-his was a gift from Gabsalilow et al. [78]. It expresses the fusion construct Strep-AvrBs4-MutH-His. AvrBs4 is truncated at the N-terminus and the C-terminus (ΔN152-C28). The linker between the repeat array of AvrBs4 and MutH consist of 28 amino acids behind the last half repeat of AvrBs4.

PZHY500 vector The shuttle vector pZHY500 was constructed by Daniel Voytas (Addgene plasmid # 36186) as compatible yeast expression vector for TALENs [246]. It harbours a truncated TALEN architecture [246] and is a Golden gate compatible expression vector [33]. The N-terminus and C-terminus are truncated (ΔN152-C68) and the repeat array is replaced by a lacZ-gene. The homodimeric FokI nuclease is fused direct to the C-terminus. pZHY500 contains a HIS3 selectable marker, as well as ampicilline resistance. For expression in yeast a TEV-promoter is used.

Cas9 system:

p414-Tef1p-cas9-cyc1t For using the CRISPR-Cas system in Saccharomyces cerevisiae the Cas9 protein was expressed from the plasmid p414-Tef1p-Cas9-Cyc1t, a gift of George Church [58] (Addgene plasmid # 43802). It contains the selective marker TRP1.

PRPR1_grna_handle_rpr1t The guided RNA was constructed starting with the plasmid pRPR1_grna_handle_RPR1t from Lu T. [69] (Addgene plasmid # 49014).

pvax This vector can be used for the expression of recombinant proteins in eukaryotic cells. It harbours a kanamycin resistance gene and expression is under control of a CMV promoter.

2.1.7.3 Substrate vectors

Pat153 was used as a substrate vector. It contains 19 unaddressed MutH recognition sites and an ampicilline resistance-gene.

Pat153_T-3-H/Pat153_T-6-H were made by Gabsalilow et al. [78] for the analysis of the in vitro activity and specificity of the AvrBs4_MutH protein. The vectors harbours the 18 bp AvrBs4 recognition site with different spacer length to the recognition site of MutH (Table 9).

Pat153_T3-4-P-4-T3 were made by Yanik et al. Yanik et al. [239] for the analysis of the in vitro activity and specificity of the AvrBs3_PvuII variants. In table 9 the inserted target site is listed.
Table 9: Addressed target sites
Specific sequences of the target site of AvrBs4 (T), AvrBs3(T3), MutH(H), and PvuII(P) are underlined. The number between T and H represents the spacer length in basepairs.

<table>
<thead>
<tr>
<th>Target site</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' - TGCAGTATAATTAATCCACTTGAGGATCCTGCTACATG-3'</td>
<td>T-3-H</td>
</tr>
<tr>
<td>5' - TGCAGTATAATTAATCCACTTGAGCGTCTGCTACATG-3'</td>
<td>T-6-H</td>
</tr>
<tr>
<td>5' - TCTATAACCTAACCCTCTATCCCAGCTGTCTAAGAGGTTAGGTTTTATAGA-3'</td>
<td>T3-4-P-4-T3</td>
</tr>
</tbody>
</table>

**PCP5 vector** from Daniel Voytas (Addgene plasmid # 35397) [33], a *E. coli*/*yeast shuttle vector, was used as a reporter plasmid in yeast. It harbours a URA3 selective marker, which is flanked by a divided lacZ-gene which contains 125 bp duplication. Additionally, it encompasses a TRP1 selective marker, kanamycin resistance and chloramphenicol resistance. The pCP5 vector uses a GDP promoter. A ccdB lethal gene, which is used as a selection marker, is also integrated in between the divided lacZ gene.

2.1.8 *E. coli* and Saccharomyces cerevisiae strains

**E. coli**

**XL10 – Gold strain** was used for the expression of pQE_Strep-AvrBs4-MutH-His and pQE_Strep-AvrBs3, as well as for several cloning steps of pZHY500.

Genotype: endA1 glv44 recA1 thi-1 gyrA96 relA1 lac Hte (mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tetR F' [proAB lacIqZΔM15 Tr10(TetR Amy CmR)]

**DH5α strain** was mainly used to sub clone the reporter plasmid pCP5.

Genotype: F- endA1 glv44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(laczYA-argF)U169, hsdR17(rK- mK+), λ−

**JM109 strain** was used for cloning the TALE binding domain by employing the Golden Gate strategy.

Genotype: endA1 glv44 thi-1 relA1 gyrA96 recA1 mcrB+ Δ(lac-proAB) e14- [F’ traD36 proAB+ lacIq lacZΔM15] hsdR17(rK-mK+)

**JM110 strain** was used to multiply the substrates inserted in pAT153. JM110 lacks methylase dam and dcm and therefore produces non methylated DNA.

Genotype: rpsL thr leu thi lacY galK galT ara tonA tsx dam dcm glv44 Δ(lac-proAB) e14- [F’ traD36 proAB+ lacIq lacZΔM15] hsdR17(rK-mK+)
Saccharomyces cerevisiae

YPH500 yeast strain was used for the yeast activity assay, which is describe in 2.2. Genotype: MATα ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ1

2.1.9 Enzymes and markers

Enzymes used were provided by Thermo Scientific Fermentas or New England Biolabs (NEB). Markers for gel electrophoresis are listed in table 10.

Table 10: Marker for gel electrophoresis

<table>
<thead>
<tr>
<th>Marker</th>
<th>Provider</th>
</tr>
</thead>
<tbody>
<tr>
<td>PageRulerTM Unstained Protein Ladder</td>
<td>Thermo Scientific Fermentas</td>
</tr>
<tr>
<td>GeneRulerTM 1kb DNA ladder</td>
<td></td>
</tr>
<tr>
<td>pUC Mix Marker, 8, ready to use</td>
<td></td>
</tr>
</tbody>
</table>

2.1.10 Primer and oligonucleotide

Cassettes used as inserts are summarized in table 11.

Table 11: Cassettes used as inserts

Specific sequences of the target site of TALE(T) and MutH(H) are underlined. The number between T and H represents the spacer length in basepairs.

Sequences of cassettes | Abbreviation
------------------------|-------------------
............5’-TCTCAGGAGGATCCCGGAAGTGGATTTAATTATACGTGACATG-3’ | H-3-T
3’-CATGAGAGTCCGCTCTAAGCCTCACAATTTAATATAGCGT-5’ |

............5’-TGCAATGATAATTATATCCACCTCTCGGATGAGATTATTTATACTGCACATG-3’ | T-3-HH-3-T
3’-CATGACGCATATTATATTAGGTGAGGCTAGGGGCTACCTAATAATATAGCGT-5’ |

............5’-TGCAATGATAATTATATCCACCTCTCGGATGAGATTATTTATACTGCACATG-3’ | T-3-H-3-T
3’-CATGACGCATATTATATTAGGTGAGGCTAGGGGCTACCTAATAATATAGCGT-5’ |

5’- CGTACGGGATACTAATATAGCT - 3’ | stop codon

............3’- CCATATGATTATCA - 5

5’- GCTGCAATGGATACATGAGCTTTAGCTAATAGCTAATAGCAAATATAGC-3’ | insert for gRNA
5’- CGACGGCTAACCCTCTGCCAAACTTTTATATATAGCGAAATCTCGTCTTATCG-3’ |
Cassettes used as target DNA for anisotropy and FRET experiments are given in table 12.

Table 12: Cassettes used as target DNA

The specific DNA contains the AvrBs3 binding site (underlined) and is either labelled with Hex at the 5' -end of the bottom strand or does not possess a fluorophore label. For fluorescence based experiments, specific DNA was labelled with Cy3 at the 5' -end of the top or bottom strand. The scramble DNA consists of the same nucleotide composition, but in a random order.

<table>
<thead>
<tr>
<th>DNA Name</th>
<th>Length [bp]</th>
<th>Fluorophores</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-TTATCTATAACCTAACCCTCTCTCT-3'</td>
<td>25</td>
<td>HEX or Cy3</td>
</tr>
<tr>
<td>3'-AAATAGATATTTGGATTTGAGAGA-5'</td>
<td>specific DNA</td>
<td></td>
</tr>
<tr>
<td>5'-ACACCTCTACCTATCTATCTTATAA-3'</td>
<td>25</td>
<td>-</td>
</tr>
<tr>
<td>3'-TGTTGAGATGGATAGAATATAGAAT-5'</td>
<td>scramble DNA</td>
<td></td>
</tr>
</tbody>
</table>

Primer used to create magaprimer for Mutagenesis-PCR are listed in table 13.

Table 13: Primers used to create Megaprimers

The first four oligos were used to create megaprimer, which induces an AvrII cutting site to the plasmid, respectively. All the rest were used to create megaprimers to substitute C (cysteine) to M (methonine) in the given plasmids.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-GAGCTCCCTAGCCAGGACGGCAAGGCGAGG-3'</td>
<td>558</td>
<td>558/455 (AvrII site in stuffer-MutH)</td>
</tr>
<tr>
<td>5'-GAGCTCCCTAGCCAGGCAAGGCGAGG-3'</td>
<td>455</td>
<td></td>
</tr>
<tr>
<td>5'-CTGAGGTCGGCCGTTGACGATGCTCTCACGACTGAC-3'</td>
<td>423</td>
<td>423/455 (AvrII site in pTal1)</td>
</tr>
<tr>
<td>5'-CTGAGGTCGGCCGTTGACGATGCTCTCACGACTGAC-3'</td>
<td>523</td>
<td>455/523 (AvrII site in pZHY500)</td>
</tr>
<tr>
<td>5'-GAGCTCCCTAGCCAGGCAAGGCGAGG-3'</td>
<td>499</td>
<td>499/498 (pHD1,8,9; pNI1,6,7; NG10)</td>
</tr>
<tr>
<td>5'-GAGCTCCCTAGCCAGGCAAGGCGAGG-3'</td>
<td>498</td>
<td></td>
</tr>
<tr>
<td>5'-GCTGTTGGCCAGTACTGACGAGGACCCCTGAGC-3'</td>
<td>501</td>
<td>501/498 (pNI3, pH03)</td>
</tr>
<tr>
<td>5'-GCTGTTGGCCAGTACTGACGAGGACCCCTGAGC-3'</td>
<td>502</td>
<td>502/498 (pNG4)</td>
</tr>
<tr>
<td>5'-GCTGTTGGCCAGTACTGACGAGGACCCCTGAGC-3'</td>
<td>503</td>
<td>503/498 (pNI5, pH05)</td>
</tr>
<tr>
<td>5'-GCTGTTGGCCAGTACTGACGAGGACCCCTGAGC-3'</td>
<td>695</td>
<td>695/498 (pH4)</td>
</tr>
<tr>
<td>5'-GCTGTTGGCCAGTACTGACGAGGACCCCTGAGC-3'</td>
<td>696</td>
<td></td>
</tr>
<tr>
<td>5'-GAGCTCCCTAGCCAGGCAAGGCGAGG-3'</td>
<td>41</td>
<td>41/686 pTal1(AvrII)</td>
</tr>
<tr>
<td>5'-GAGCTCCCTAGCCAGGCAAGGCGAGG-3'</td>
<td>696</td>
<td></td>
</tr>
</tbody>
</table>
PRIMER used for screening and sequencing are shown in table 14.

Table 14: Primers
Oligos used for screening

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’-GTATACGAGGCCCTTTCGTCT-3’</td>
<td>126 (for)</td>
</tr>
<tr>
<td>5’-CATTTACTGGATCTATCAACAGGAG -3’</td>
<td>68 (rev)</td>
</tr>
<tr>
<td>5’-catcgccaatgcactgac-3’</td>
<td>455 (for)</td>
</tr>
<tr>
<td>5’-ggcgacaggtgtgctgtg-3’</td>
<td>454 (rev)</td>
</tr>
<tr>
<td>5’-tggcgctggcaaaacagtgg -3’</td>
<td>453 (for)</td>
</tr>
<tr>
<td>5’-GGAGAAAATACCCGATCAG -3’</td>
<td>477 (rev)</td>
</tr>
<tr>
<td>5’-CGGTTGCCACCCCTGTCTCCCTTG-3’</td>
<td>MutHseq4 (rev)</td>
</tr>
<tr>
<td>5’-TCATCGCAAGACGCGCAACAGGAT -3’</td>
<td>577 (rev)</td>
</tr>
<tr>
<td>5’-CAACCTTCCACCATTGCCTGAAAAATATCATTTA -3’</td>
<td>457 (rev)</td>
</tr>
<tr>
<td>5’-TCCCCAAGCTACTCATGAGCTATTGTTG -3’</td>
<td>564 (rev)</td>
</tr>
<tr>
<td>5’-ACTCTGCCACCATGCTCTGCT -3’</td>
<td>476 (rev)</td>
</tr>
<tr>
<td>5’-ATGCACCTAGATCTGGGACCCACCCCTTCAG -3’</td>
<td>473 (for)</td>
</tr>
<tr>
<td>5’-CCTCAGCTACTGAGCCCGTTGAGCACCACCGC -3’</td>
<td>474 (rev)</td>
</tr>
<tr>
<td>5’-TTTATCTTATATATCTCTTGAGACTCCACGCTCTGTAGACATG -3’</td>
<td>339 (for)</td>
</tr>
<tr>
<td>5’-GGATCCACAGGTCTTAGAAGGGTTATATATGATTATGCCAATG -3’</td>
<td>337 (rev)</td>
</tr>
<tr>
<td>5’-TGACCTCCCATTGATTTAACCTTAAACG -3’</td>
<td>487 (for)</td>
</tr>
<tr>
<td>5’-GTTTCGGGAGCATGCAGCCGGCTTGG-3’</td>
<td>lab S/H (for)</td>
</tr>
<tr>
<td>5’-GCACCGCGCGCGCAAGGAATGGTG-3’</td>
<td>t4-x-H (rev)</td>
</tr>
</tbody>
</table>
2.2 Methods

Microbiological standard methods (producing competent cells, transformation of plasmids in *E. coli* cells or *Saccharomyces cerevisiae* cells etc.) and microbiological standard methods (agorse- and polyacrylamide-electrophoresis, DNA isolation, restriction digest etc.) were performed as described in *Current Protocols in Molecular Biology* and *Protein Science* [10, 46].

2.2.1 Molecular cloning

Molecular cloning is a technique to assemble recombinant DNA, which is then replicated in a host organism. Recombinant DNA was introduced to the host bacterium *E. coli*. In figure 8 the basic principles of molecular cloning are shown. First, an insert has to be created, which is then ligated into a plasmid. Inserts can be constructed by cassettes, PCR amplification or digestion. Cassettes used as inserts for cloning strategies are listed in table 11. These oligos were mixed in a 1:1 ratio, heated at 95°C for 3 minutes and then slowly cooled down to RT. The remaining inserts were processed either by PCR amplification of the gene of interest or by digestion of a plasmid containing the gene of interest. The amplified PCR fragments were digested by the required restriction enzymes. Receiving the insert through digestion, the DNA-fragments were separated by gel electrophoresis and the fragment were purified by the Wizard R SV Gel and PCR Clean–up System (Promega). The needed backbones were prepared by digesting the plasmid with restriction enzymes and dephosphorylation of the DNA ends by Fast AP (37°C, 15 min.). Digestion of DNA fragments or plasmids were preformed at 37°C for 1 hour. For ligation 5 nM insert, 3 nM backbone DNA, 1 x T4-ligase buffer and 1 μl T4–ligase, as suggested by the manufacture, were mixed in 20 μl. Ligation were performed at 22°C for 1 h or 16°C overnight. Z–competent cells were prepared for transformation by the Z-competent *E. coli* Transformation Kit & Buffer Set (Zymo Research) and stored at -70°C. For transformation cells were allowed to defrosted for 10 minutes on ice and 3-5 μl ligation–mixture was added to the cell aliquot. Further, the cell aliquots stayed on ice over 5–10 minutes and 400 μl SOC media was added. Cells were permitted to recover for about an hour at 37°C before distributing the cell aliquot on the essential LB–plate (37°C, o/n). Figure 8 depicts the key steps of the cloning strategies.

**Mutagenesis-PCR** [118] was used to substitute precise amino acids. First, a mega primer was amplified by PCR using primers including the desired mutation. Next, the mega primer was then used for a rolling circle PCR (Table 16), which amplified the complete vector, containing the new mutation.
Figure 8: Basic steps of the cloning strategies
(A) Insert made by PCR amplification, followed by double digestion or by immediate double digestion of the plasmid. (B) The backbone was obtained by double digestion of the starting plasmid. The desired recombinant plasmid was gained by ligation.

Table 15: Components used for PCR are listed.

<table>
<thead>
<tr>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 ng DNA-matrix</td>
</tr>
<tr>
<td>400 nM 5’-primer</td>
</tr>
<tr>
<td>400 nM 3’-primer</td>
</tr>
<tr>
<td>200 μM dNTPs</td>
</tr>
<tr>
<td>1x Pfu-buffer</td>
</tr>
<tr>
<td>1 μl Pfu-polymerase</td>
</tr>
</tbody>
</table>

For exact amplification the pfu–polymerase of Pyrococcus furiosus (3’- 5’ Exonuclease activity, \( T_{\text{opt.}} = 68^\circ \text{C} \), \( \sim 500 \) nucleotide per minutes) was used. To digest the template vector the mixture was digested by DpnI, an enzyme which only cuts at methylated sites.

**Blue-white screening** Blue-white screening was used to simplify the cloning procedure, but the same principle was also applied in the single strand annealing assay, as described later. The method is based on an enzyme, called \( \beta \)-galactosidase. This enzyme is able to convert 5- bromo-4-chloro-3-indolyl-\( \beta \)-D-galactopyranoside (X–Gal) to a blue dye (figure 9). \( \beta \)-galactosidase is encoded by the lacZ gene of the lac operon and functions as a homotetramer. *E. coli* strains carrying the lacZ deletion mutant (lacZ\( \Delta \)M15) express a \( \beta \)-galactosidase (\( \omega \)-peptide), which misses its N-terminal residues, resulting in a non functional enzyme. If the *E. coli* strain carries a plasmid, which encodes the first 59 residues (\( \alpha \)-peptide) of \( \beta \)-galactosidase the \( \omega \)-peptide and the \( \alpha \)-peptide reconstitute a functional \( \beta \)-galactosidase enzyme.
Table 16: PCR-programs

PCR-programs applied for this thesis are described. The basic program is used to amplify DNA fragments. T_{annealing} depend on length and composition of the primer and T_{opt} depends on the employed polymerase (e.g. Taq-polymerase T_{opt.} = 72°C). The rolling circle program is used for mutagenesis-PCR. Prior to that, a megaprimer is processed, which contains the relevant mutation.

<table>
<thead>
<tr>
<th>basic program</th>
<th>rolling circle program</th>
</tr>
</thead>
<tbody>
<tr>
<td>temperature</td>
<td>time</td>
</tr>
<tr>
<td>[C°]</td>
<td>[min]</td>
</tr>
<tr>
<td>94</td>
<td>2</td>
</tr>
<tr>
<td>92</td>
<td>0.75</td>
</tr>
<tr>
<td>T_{annealing}~</td>
<td>0.50</td>
</tr>
<tr>
<td>58-62</td>
<td></td>
</tr>
<tr>
<td>T_{opt.}</td>
<td>500-1000</td>
</tr>
<tr>
<td>nt/min</td>
<td>go to step one and</td>
</tr>
<tr>
<td></td>
<td>repeat 30 x</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>T_{opt.}</td>
<td>4</td>
</tr>
<tr>
<td>break</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>go to step one and</td>
</tr>
<tr>
<td></td>
<td>repeat 20 x</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td>break</td>
<td>20</td>
</tr>
</tbody>
</table>

The LB-plates have to be spread with 40 µl 20 mg/ml X-Gal and 40 µl of a 0.1 M Isopropyl β-D-1-thiogalactopyranoside (IPTG). IPTG is needed to inactivate the LacZ repressor in E. coli. The colonies growing on the plate will either be white or blue, depending on the existence of the α-peptide encoding plasmid.

Figure 9: X-Gal converted to a blue dye

β-galactosidase hydrolyse X-Gal to galactose and 5-bromo-4-chloro-3-hydroxyindole. Due to spontaneous dimerisation followed by oxidation, a 5,5′-dibromo-4,4′-dichloro-indigo is yielded and responsible for the blue color.
2.2.1.1 **Golden Gate**

The TALE binding domains were designed with the **Golden Gate** TALEN and TAL Effector Kit 2.0 from Daniel Voytas and Adam Bogdanove (Addgene Kit#1000000024) [33]. The kit consists of bacterial glycerol stocks enclosing the required plasmids to construct a TALE binding domain. The kit enables to design TALE binding domains enclosing 12-31 repeats. The kit contains various plasmids. The backbone scaffold (pTAL1-4) contains the TAL-effector gene. Here the repeat array is replaced with a lacZ gene, which functions as stuffer. The lacZ gene enabled blue-white screening.

Plasmids containing the encoding sequence for one repeat (e.g. pNG1) and a tetracycline resistance and plasmids carrying the encoding sequence for a last half repeat (e.g. pLR-HD) with a spectinomycin resistance are included. Plasmids named, pFusA(1-10) or pFusB(1-10) can assemble up to 10 repeats in one plasmid and enclose a spectinomycin resistance-gene.

The principle of **Golden Gate** cloning strategy is to use typeII restriction enzymes, which induce DSBs outside of their recognition sites. Due to that, sticky ends emerge which can exclusively be ligated to a compatible other sticky end. The kit enables to assemble ten repeats into one backbone in a one pot reaction. In figure 10, the principle is illustrated. The protocol of the **Golden Gate** TALEN and TAL Effector Kit 2.0 take five days to construct one binding domain and we followed the protocol exactly as recommended by the manufactures.

All constructed TALE binding arrays were assembled into a modified pTAL1 vector, which contained an AvrII cutting site. This enabled us to insert the TALE binding arrays into the specific expression vector. We used modified versions of pZHY500[246] for yeast expression, pVAX for expression in HEK297 cells and pQE-30 vectors for expression in *E. coli*. In table 17 the repeat composition of the designed TALE binding domains are given.

2.2.2 **vectors**

**Expression vectors** The expression vectors were modified for the insertion of the TALE binding domains. An AvrII cutting site was introduced by **Mutagenesis-PCR** to pTAL1 and pZHY500. This enabled us to remove the TALE binding domain from the pTAL1 vector with the restriction enzymes AvrII and Stul and to insert the repeat array into the modified pZHY500 expression vector.

**Yeast expression vector pZHY500_AVRBS4-FOKI**: The sequence encoding AvrBs4 was inserted to create the coding sequence for the homodimeric AvrBs4-FokI.
Table 17: Repeat composition assembled with the Golden Gate cloning strategy

Two naturally occurring TALE binding domains were designed (AvrBs4 and AvrBs3). Additionally, five binding domains addressing a GATC site (I) and (II) in the RPGR-gene were constructed.

<table>
<thead>
<tr>
<th>Repeat composition</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>NI1 NG2 N13 N14 NG5 NG6 NI7 N18 N9 N10 NI1 N62 HD3 HD4 N15 HD6 N77 N6</td>
<td>AvrBs4</td>
</tr>
<tr>
<td>HD1 NG2 NS5 NG4 NI5 NI6 NI7 HD8 HD9 NS1 NS2 HD3 HD4 HD5 NG6 HD7 NG</td>
<td>AvrBs3</td>
</tr>
<tr>
<td>TALE(RPGR)</td>
<td></td>
</tr>
<tr>
<td>HD1 NI2 NH3 NH4 NS5 NH6 NH7 NH8 NH9 NI10 NI2 NH3 NI4 NS5 NI6 NH7 NI</td>
<td>1(I)</td>
</tr>
<tr>
<td>NG1 NG2 NH3 NI4 NG5 HD6 HD7 HD8 HD9 NH10 NG1 HD2 NI3 HD4 HD5 HD6</td>
<td>2(I)</td>
</tr>
<tr>
<td>HD1 NI2 HD3 NI4 NH5 NS6 N77 NS8 NS9 NH10 NI1 NI2 HD3 HD4 HD5 NH6 N77 HD8 NO</td>
<td>3(I)</td>
</tr>
<tr>
<td>HD1 NG2 N13 NI4 N5 HD6 NG7 HD8 NI9 HD10 HD1 HD2 NG3 HD4 NG5 HD6 NI7 HD8 N19 NN10 NI</td>
<td>1(II)</td>
</tr>
<tr>
<td>NH1 NG2 NG3 NG4 HD5 NG6 N17 N18 N19 HD10 NG1 NH2 NI3 HD4 HD5 NG6 NG7 HD8 NG9 HD10 NI1 HD</td>
<td>2(II)</td>
</tr>
</tbody>
</table>
**PZHY500_AVRBS4**: A negative control vector (AvrBs4-stop) was designed by replacing the homodimeric FokI cleavage domain by the stop codon cassette (5'-CTAGGCGGTACCTAATAGTAGCT-3', 3'-GATCCGCCATGGATTATCATCGA-5') inserted in the AvrII and HindIII sites.

**PZHY500_I-SCIE**: The plasmid coding for I-SceI was created by amplifying the I-SceI gene using pASK_His-MutH-I-SceI-Strep as a template and inserting the PCR fragment into the vector pZHY500 following he cutting with XbaI and SacI.

**PZHY500_AVRBS4-MUTH**: To construct an analogue yeast expression vector for AvrBs4-MutH the lacZ-gene was amplified by PCR and inserted by StuI and SacI as a stuffer replacing TALE in the *E. coli* expression vector pQE30_Strep-TALE-MutH-His6 [76]. Next, stuffer-MutH was amplified and integrated into pZHY500 using StuI and Ecl136, resulting in pZHY500_stuffer-MutH. An AvrII site was introduced by mutagenesis PCR and the stuffer (lacZ-gene) was replaced by the encoding sequence of AvrBs4.

**PZHY500_AVRBS3-PVUII (VARIANTS)**: To construct analogue yeast expression vectors for AvrBs3-PvuII variants, the MutH-encoding gene in pZHY500_lacZ-MutH was replaced by the respective PvuII (hfPvuII, scPvuII, D30KPvuII, and K38DPvuII) encoding gene using AvrII and Ecl136 sites. The stuffer (lacZ-gene) was replaced by the encoding sequence of AvrBs3 using the AvrII and the StuI site.

**HEK297 EXPRESSION VECTOR PVAX_TALE(RPGR)**: All TALE arrays addressing the RPGR-gene were inserted into pVAX_stuffer-MutH. The encoding sequence stuffer-MutH was gained as described above and inserted into the pVAX vector using StuI and Ecl136 sites. The required pVAX-TALE(RPGR) vectors were constructed by using the AvrII and StuI cutting sites.

*Cysteine to methionine substitution in AvrBs3*

Additionally, different variants of the TAL effector AvrBs3 were designed using the plasmid kit for the generation of TALENs as described by Daniel Voytas and Adam Bogdanove (Addgene kit # 1000000024) [33]. Cysteine residues at position 30 in all single 17 repeats were substituted by MUTAGENESIS PCR to methionine. Furthermore, the cysteine in the backbone pTAL1 was removed by mutagenesis PCR. An AvrII cutting site was introduced into the backbone pTAL1, to relocate AvrBs3 variants into an *E. coli* expression vector pQE30_Strep-AvrBs3.
Table 18: AvrBs3 variants

TALEs made with Golden Gate strategy. TALEs were cloned into pTAL1 + AvrII backbone. RVDs which are underlined include cysteine to methionine substitution. The last half repeat contains no cysteine. (I) represent AvrBs3 without any cysteine; (A) represent AvrBs3 (A) carry one cysteine residue in repeat domain one; (B) AvrBs3 (B) enclose one cysteine residue in repeat domain sixteen; (AB) AvrBs3 (AB) contain one cysteine residue in repeat domain one and one cysteine domain in repeat number sixteen.

<table>
<thead>
<tr>
<th>repeat composition</th>
<th>name</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD1 NG2 NS3 NG4 N15 N16 N17 HD8 HD9 NG10 NS1 N52 HD3 HD4 HD5 NG6 HD7 NG</td>
<td>(I)</td>
</tr>
<tr>
<td>HD1 NG2 NS3 NG4 N15 N16 N17 HD8 HD9 NG10 NS1 N52 HD3 HD4 HD5 NG6 HD7 NG</td>
<td>(A)</td>
</tr>
<tr>
<td>HD1 NG2 NS3 NG4 N15 N16 N17 HD8 HD9 NG10 NS1 N52 HD3 HD4 HD5 NG6 HD7 NG</td>
<td>(B)</td>
</tr>
<tr>
<td>HD1 NG2 NS3 NG4 N15 N16 N17 HD8 HD9 NG10 NS1 N52 HD3 HD4 HD5 NG6 HD7 NG</td>
<td>(AB)</td>
</tr>
</tbody>
</table>

CRISPR-CAS SYSTEM: The guided RNA was constructed starting with the plasmid pRPR1_gRNA_handle_RPR1t from Lu T. [69] (Addgene plasmid # 49014). Using a one-step Gibson assembly reaction a 20 bp reporter (5’-GGATAACAGGGTAATATAGC-3’) was introduced into the gRNA-expression plasmid. The target site addressed overlap with the I-SceI target site.

GIBSON ASSEMBLY The Gibson Assembly® Cloning Kit from NEB was used to design the gRNA. The HindIII site in pRPR1_gRNA_handle_RPR1t was employed to insert the specific sequence via an one-step Gibson assembly. The cassette used for insertion carried the 20 bp homologous sequence upstream and downstream of the HindIII site. In between the 20 bp duplications sequence, the 20 bp sequence addressing the Cas9 target site is located. HindIII digestion was followed by 5′-end resection, the homologous parts anneal with each other and the DNA polymerase extends the 3′-ends. Finally, the DNA ligase seals the nicks. The procedure was performed as recommended by the manufactures.

SUBSTRATES The substrate plasmids pat153_T3-H-3-T, pat153_T3-HH-3-T and pat153_H-3-T were constructed as has been previously described by Gabsalilow et al. [78]. The appropriate DNA-cassette enclosing the target site was inserted to pat153 by using the KpnI and SphI cutting sites.

To design and construct the E. coli/Yeast shuttle vectors the target sites of the different pat153 plasmids were amplified and inserted via the BgIII and SpeI site into pCP5, a kind gift from Daniel Voytas [33] (Addgene plasmid # 35397).

The target site (pCP5_T3-4-P-4-T3) for the AvrBs3-PvuII variants was constructed similarly by Dr. Mert Yanik.
2.2.3 Protein expression and purification

AvrBs3-MutH

Strep-AvrBs4(28)–MutH-His (Figure 13), containing a Strep-tag and a His-tag, was purified by the tandem purification, recently described by Gabusallilow et al. [78]. E. coli strain XL-10 Gold was transformed with the expression plasmid. 25 ml LB-media containing ampicillin (75 µg/ml) was inoculated and incubated overnight at 37°C. Next, 500 ml LB-media, containing ampicillin, was inoculated with the overnight culture and shaken at 37°C until an optical density (OD600) of 0.7 -0.9 was reached. Expression was induced with 1 mM isopropyl-β-d-thiogalactopyranoside (IPTG). The LB-media was incubated overnight at 20°C. The cell suspension was centrifuged (3600 × g, 4°C; 15 min). The supernatant was removed and cells were washed with 40 ml His-binding-/ washing buffer (20mM tris-HCl, 1 M NaCl, 20 mM imidazole, 1mM phenylmethanesulfonyl fluoride (PMSF), pH 7.9). Cell lysis was achieved by sonification (Duty Cycle 50 %, output Control 5; 12 x 15 sec) and cell extracts were centrifuged (31000 × g, 30 min at 4°C). Meanwhile Ni-NTA agarose (Qiagen) was equilibrated with 30 ml His - binding -/ washing buffer for 30 min at 4°C. The equilibrated Ni–NTA was incubated with the clean cell extract for 1 h at 4°C and washed twice afterwards with 40 ml His-binding/- washing buffer. The Ni-NTA beads were transferred to a Poly–Prep chromatography column (Poly-Prep, BioRad)) and the protein was eluted with 3 x 1 ml elution buffer (20mM tris-HCl, 1 M NaCl, 200 mM imidazole, pH 7.9). The eluates of the His–purification were loaded on a column with Strep–Tactin sepharose (IBA). The Strep-Tactin spher-rose was washed with 4 ml Strep-binding/- washing buffer (100 mM tris-HCl, 1 M NaCl, pH 7.9) and the protein was eluted with 5 x 500 µl Strep–elution buffer (100 mM tris-HCl, 500 mM NaCl, 2.5 mM, desthiobiotin, pH 7.9). The eluates were combined and concentrated at 4°C by dialysis into 10 mM HEPES-KOH, pH 7.9, 500 mM KCl, 1 mM EDTA, 1 mM dithiothreitol (DTT) and 50% (v/v) glycerol. The protein was stored at -20°C and concentrations were determined by absorbance (280 nm) according to Pace et al. [162].

AvrBs3 variants

Strep-AvrBs3 variants (Figure 33), containing a Strep-tag were purified using a Strep-Tactin sepharose column as describe in chapter 2.2.3. Except that the supernatant after cell lysis was filtered with a syringe filter (0.45 µm Celluloseacetat Membran, VWR) and loaded on the Strep–Tactin sepharose column. AvrBs3 variants were labeled in the next step or purified in a second step over heparin sepharose columns. The protein eluate of the Strep-tag purification was diluted to a salt concentration of 160 mM. The ion exchange chromatography was performed with Äktapurifier 10 Chromatographic System (GE
Specific nuclease induces a DSB and triggers SSA. The blue arrows indicate the partially duplicated fragments of the lacZ gene. DSB is followed by 5’-end resection and the duplicated parts anneal to each other. This results in overhangs, which are resected and the gaps are filled with nucleotides. Finally, the junction is ligated and resulting in a functionally lacZ gene.

Healthcare). The gradient was shaped by a high (1000 mM NaCl) and low salt (150 mM NaCl) buffer and AvrBs3 variants eluted at a NaCl concentration of 285 nM. The eluate were combined and concentrated at 4°C by dialysis into 10 mM HEPES-KOH, pH 7.9, 500 mM KCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM dithiothreitol (DTT), 50% (v/v) glycerol.

2.2.4 Eukaryote cell systems

Single strand annealing assay

To determine the specific in vivo activity and efficiency of the analyzed nucleases, a yeast based recombinant assay was established according to the method previously described by Cermak et al. [33]. Briefly, successful cleavage of the target positioned between two partially duplicated fragments of the lacZ gene reconstitutes a functional lacZ gene due to subsequent single strand annealing (figure 11), thus enabling a quantitative analysis of the nuclease activity. Yeast cells YPH500 were co-transformed with a 1:1 ratio of the reporter plasmid (pCp5_addressed site) and the expression plasmid (pZHY500_specific nuclease). Three different amounts of respective inserted DNA were tested (1000 ng, 2000 ng, and 3000 ng). 200 µl of the transformation-mixture were evenly dispersed on a drop out media plate, containing 0.08 mg/ml 5’-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-GAL). After 5 days of incubation at 30°C, outgrowth of blue and white colonies was counted. All designed nucleases were compared to I-SceI on target T-3-HH-3-T. Levels of significance were determined by unpaired T-tests (P < 0.05 *, P < 0.01 **, P < 0.001 ***).
2.2.4.1 HEK293T cytotoxicity assay

The TALE-MutH constructs were analyzed in terms of their cytotoxicity and activity in HEK293T cells together with Dr. Yanik in the retinal gene therapy lab of Prof. Dr. Knut Stieger.

**Cytotoxicity assay** [49]: To determine the nuclease-associated toxicity, HEK293T cells were seeded in a 24-well plate at a density of 150,000 cells per well in Dulbecco’s modified Eagle media supplemented with 10% fetal bovine serum and penicillin/streptomycin (Invitrogen). The cells were transfected using polyethylenimin (PEI)-mediated transfection [203] containing the TALE-MutH pair TALE-RPGR-1(I)-MutH (600 ng)/ TALE-RPGR-2 (I)-MutH (600ng) expression vectors and the eGFP expression vector (100ng, eGFP-N1, Clontech). For comparison, the toxic ZFNs GZF1N/GZF3N [6] (kindly provided by T. Cathomen, University Medical Center Freiburg, Germany), the **golden standard** nuclease I-SceI (I-SceI was cloned into the pvax plasmid, Invitrogen) and an empty vector pcDNA3.1 (Invitrogen) were analyzed similar. The number of eGFP-positive cells were determined by flow cytometry (FACSCalibur; BD Biosciences) at day 2 and 4 after the transfection. The survival rate is analyzed by counting the decrease of eGFP-positive cells in this interval and is normalized to I-SceI.

**Traffic Light Reporter (TLR) system:** The traffic light reporter system has been described by Certo et al. [34] and was employed to determine the activity of the TALE-MutH pair (TALE-RPGR-1(I)-MutH/ TALE_RPGR-2 (I))-MutH). The TLR system enclosing a I-SceI recognition site was modified by replacing the incorrect mCherry encoding sequence to the incorrect blue (BFP) encoding sequence using the InFusion cloning (Clontech) system. NHEJ, triggered by I-SceI, causes a frame shift which result in the correct BFP reading frame. BFP positive cells were counted by flow cytometry (FACSCalibur; BD Biosciences, kindly provided by H. Hackstein, JLU Gießen). The recognition site for the TALE-MutH pair was introduced adjacent to the I-SceI target site using a AvrII and BsiWI restriction site.

HEK293T cells were seeded similarly as for the cytotoxicity assay. After 24 h, cells were transfected with the PEI transfection mix (PEI (0.1 g/l, pH: 5.5, 150 mM NaCl) containing 600 ng of the TLR plasmid enclosing the specific target site and 600 ng of the specific nuclease expression vector. Cells were harvested at day 3 after the transfection and the number of BFP positive cells was determined by flow cytometry (FACSCalibur; BD Biosciences).
2.2.5 Cleavage assays

To determine the cleavage specificity and activity in vitro two cleavage assays, recently described by Gabsalilow et al. [78] were used.

Cleavage assay on plasmid

The in vitro activity of Strep-AvrBs4(28)-MutH-His6 on H-3-T and T-3-H were analyzed on a plasmid containing the addressed site and 18 unaddressed GATC sites. 8 nM substrate were incubated with 32 nM Strep-AvrBs4(28)-MutH-His6 at 37°C (20 mM Tris-acetat, 120 mM K-acetat, 1 mM MgCl2, 50 µg/ml bovine serum albumin (BSA), pH 7.5). Aliquots of the reaction were stopped with 50 mM EDTA at defined time intervals (1', 3', 10', 30', 60', 120', 180') and analyzed by an 1.2% agarose gel electrophoresis. The gels were stained by ethidium bromide and analyzed using LabImage 1D.

Cleavage assay on fluorescently labelled PCR fragments

To determine the strand-specificity of the substrates fluorescently labelled PCR products were used. To amplify the PCR products (T-3-H, H-3-T, T-6-H, T-3-HH-3-T and T-3-H-3-T), 5’-end labeled (Atto 488 or Atto 647N) forward and reverse primers were used. 20 nM labeled PCR-substrate were incubated with six-fold excess of enzyme at 37°C (20 mM Tris-acetat, 120 mM K-acetat, 1 mM MgCl2, 50 µg/ml (BSA), pH 7.5). Reaction were stopped by 50 mM EDTA at defined time points (3’, 10’, 30’, 60’, 120’, 180’) and analyzed by denaturingpolyacrylamide gel electrophoresis (10% polyacrylamide 19:1 (PAA), 7M urea, formamide 30%, 1 x TBE).

Similarly, we incubated the targets with AvrBs4-MutH for 180 min, followed by digestion with 1 µl Proteinase K (10 min at 37°C), followed by protein heat deactivation for 10 min at 95°C. New 120 nM AvrBs4-MutH was added and allowed to digest the substrate for another 60 min at 37°C. The defined reaction steps were stopped with 50 mM EDTA and analyzed by denaturing PAGE (heat: 400 V, 40 mA, 30 min; run: 300 V, 30 mA, 45 min) and native 10% (PAA) gels (run: 300 V, 30 mA, 45 min). The fluorescent images were visualized with TyphoonTM FLA 9500. Alexa 488 was excited at 473 nm using the BPG1 (560-580 nm) filter and Atto647N was excited at 635 nm using the LPR (> 665 nm) filter.
2.2.6  Fluorescence based methods

Labeling of AvrBs3 variants

The proteins were all labeled by a thiol Michael-type 1,4-addition reaction, creating a covalent connection between the thiol group of the cysteine residue and the maleimide fluorophores. The reaction mechanism is shown in figure 12. The thiol group is deprotonated in presence of a base. Next, the thiol group is added to the double bond of the maleimide creating an intermediate (I). Hydrolysis of the intermediate result in the succinimide-thiolether (II).

![Reaction Mechanism](image)

Figure 12: Thiol Michael-type addition reaction
(A) Equilibrium reaction of the thiol group of the cysteine with the maleimide fluorophore. (B) Reaction mechanism of the thiol Michael-type 1,4-addition. The thiol group is deprotonated in presence of a base and is added to the maleimide by creating an intermediate (I). Hydrolysis results in the succinimide-thiolether (II).

Labeling conditions: 5 mM redox reagent Dithiothreitol (DTT) was added to 80 µM of the protein and incubated at 0°C for 1 h. Dialyse puffer was exchanged to an adjusted puffer for labeling (10 mM 2-(4-(2-hydroxyethyl)-1-piperazinyl)-ethansulfonsäure (HEPES) /KOH pH 8.0, 1 mM EDTA, 200 mM KCl, 10% glycerol) by a zeba spin desalting column 40 K. The fourfold excess of the maleimide dye was supplemented to the single cysteine AvrBs3 variants and the twofold excess of maleimide dye was added to the AvrBs3 (AB) variant. The
protein was incubated at 0°C under exclusion of light for 1-2 hours. AvrBs3 variants were purified by the zeba spin desalting column 40 K and eluted twice with the high salt labeling puffer (10 mM 2-(4-(2-hydroxyethyl)-1-piperazinyl)ethansulfonsäure (HEPES)/KOH pH 8.0, 1 mM ethylenediaminetetraacetic acid (EDTA), 500 mM KCl, 10% glycerol). The aliquots were frozen with liquid nitrogen and stored at -80°C.

2.2.6.1 Degree of labeling (DOL)

The degree of labeling was determined after each labeling reaction of a protein with a fluorophore. DOL describes the ratio between the dye and the protein concentration (Formula (1)).

\[
DOL = \frac{c(dye)}{c(protein)}
\]  

(1)

The concentrations of the dye and the protein were measured with the spectrophotometer NanoDrop® 1000 (Thermo Scientific).

\[
c(dye) = \frac{A_{max}}{\varepsilon_{max} * d}
\]  

(2)

\[
c(protein) = \frac{A_{prot}}{\varepsilon_{prot} * d}
\]  

(3)

\[
A_{prot} = A_{280nm} - A_{max} * c_{f280}
\]  

(4)

\[
c_{f280} = \frac{\varepsilon_{max}}{\varepsilon_{280nm}}
\]  

(5)

Similar to the determination of the protein concentration the DNA concentration was analyzed and could be used to define the DOLs of the labeled DNAs. The absorption maximum of protein exist at 280 nm, whereas the absorption maximum of DNA is at 260 nm. Thus, the absorption at 260 nm (\(A_{260\text{ nm}}\)) was measured and the extinction coefficient \(\varepsilon_{260\text{nm}}\) was required to calculate the DNA concentration.

2.2.6.2 Electrophoretic Mobility Shift Assay (EMSA)

EMSA can be used to analyze DNA-protein or RNA-protein interaction. We applied an EMSA based on a fluorescently labeled DNA. A 25 bp long DNA containing the AvrBs3 recognition site labeled with Cy3 at the 5’-end of the lower strand was used. The unbound DNA contains a high mobility due to its small size. Binding of protein to
2.2 METHODS

the DNA results in a reduction in the mobility and can be analyzed on a gel. 5′-end Cy3 labeled DNA (20 nM) was preincubated under high salt conditions (150 mM) in 1x NEB 3.1 buffer with protein (0-600 nM) for 5 minutes on ice. The reaction was stopped with 50 mM EDTA and the whole reaction was loaded to a 4% polyacrylamid gel (35 mA, 100 V, 20 min, 4°C). The fluorophore Cy3 (filter: BPG(560-580 nm)) and Alexa 594 (filter: LPG (>575 nm)) was excited the the green SHG 532 nm, the fluorophore Alexa488 (filter: BPB(520-540 nm)) was excited with the blue LD laser (473 nm), Alexa 647 (filter: LPR (>665 nm)) was excited with the red LD laser (635 nm) and the gel was visualized with the TyphoonTM FLA 9500.

Fluorescence anisotropy

Fluorescence anisotropy is a tool to analyze binding constants and kinetics. It depends on the rotation time of the fluorophore part or the whole molecule. An advantage of fluorescence anisotropy is that only one molecule has to be labeled. For instance, using fluorescence anisotropy to investigate DNA::protein binding, the DNA can carry the fluorophore or the protein. Therefore, the protein does not has to be modified, thus excluding the risk of changing the binding character of the analyzed protein.

The fluorophore is excited with linear polarized light. It absorbs a photon, which results in a higher energy level of the valence electrons. The fluorescence lifetime $\tau$ describe the timephase the higher energy level exist. The energy is mostly released in parts by oscillation and the remaining energy is emitted by a photon. The fluorophore molecule has to be oriented in the angle of the polarized light to absorb the photon. If the molecule tumbling is “slow” the emitted photon will be polarized similar to the applied polarized light. whereas, a small fluorophore molecule will change the orientation before emitting the photon, thus resulting in a different orientation. Taken together, fluorescence anisotropy is dependent on the rotation lifetime $\phi$ and the fluorescence lifetime $\tau$ of the fluorophore. To detect this effect, the fluorescence intensity is measured parallel and perpendicular to the polarized light by aligning the polariser and the analyzer to each other [123].

Binding characteristics of the AvrBs3 variants to the DNA were analyzed by fluorescence anisotropy. The assays were performed under low salt conditions (60 mM) in assay puffer (20 mM tris-acetat, 50 mM kkalium-acetat, 10 mM magnesium-acetat, 1 mM DTT, pH 7.9) and fluorescence anisotropy was measured by using the Fluoromax-4 Spectrofluorometer to determine the $K_D$ value. The 25 bp long DNA contained the AvrBs3 binding site and was labeled with Hex at the 5´ prime end of the bottom strand. To determine the $K_D$-value 0.5 nM of the DNA was set constant and the protein was titrated to the solution.
To define the active rate of protein, the titration was repeated with 10 nM DNA. The kinetic experiments were performed under the same buffer conditions, but using the Tecan infinitie F200 pro plate reader. 10 nM labelled DNA with 50 nM AvrBs3 was used to achieve complete binding. For competition experiments unlabeled specific DNA (180 nM) and scramble DNA (180 nM) were used.

**Förster-Resonanzenenergietransfer (FRET)**

FRET is a physical process, which enable to measure the acceptor emission resulted by donor excitation. Thus, FRET requires a donor fluorophore and an acceptor fluorophore, which can be labeled to on protein (intramolecular) or to diverse biomolecules (intermolecular). The radiation-free energy transfer occur, when defined requirements are given. The donor emission spectra have to overlap with the acceptor excitation spectra (I); the orientation of the transient dipole moments have to bee favourable (II); and the proximity between donor and acceptor fluorophore has to bee given (1-10 nm) [74, 75, 73, 244].

2.2.6.3 **FRET experiment between Cy3 and Cy5**

The FRET experiments were performed under low salt conditions (80 mM) with assay puffer (20 mM tris-acetat, 50 mM kalium-acetate, 10 mM magnesium-acetat, 1 mM DTT, pH 7.9). The specific DNA (Table 12) was labeled with Cy3 at the 5’-end of the bottom (specific DNA (I)) or the top (specific DNA (II)) strand. AvrBs3 (A)/AvrBs3 (B) labeled with Cy5 was titrated to 1 nM of the labelled specific DNA, respectively.

To calculate the FRET index \(\frac{nF}{F_{D}^{\text{ex}}F_{A}^{\text{em}}[\%]}\) the three cube method described by Youvan et al. [242] was used (Formula 6). This method describes no FRET efficiency, but a values, which change with FRET [244]. Three measurements were required, when donor and acceptor fluorophores were presence. First, the donor was excited and emission in acceptor \(F_{A}^{\text{ex}}F_{A}^{\text{em}}\) and donor \(F_{D}^{\text{ex}}F_{D}^{\text{em}}\) channel was detected. Next, the acceptor was excited and emission was measured in the acceptor channel, resulting in the \(F_{A}^{\text{ex}}A_{A}^{\text{em}}\) signal.

The donor fluorophore Cy3 was excited at 520 ± 5 nm and the acceptor fluorophore Cy5 was excited at 650 ± 5 nm using the Fluoromax-4 Spectrofluorometer. To consider crosstalk of Cy3 and Cy5 both Cy3 labelled DNA and Cy5 labelled AvrBs3 were excited individually, and the resulting emissions in the donor channel and in the acceptor channel were measured (Formula 7 and 8). FRET efficiency was calculated with the following formulas.
\[ nF = F^{exD,emA} - (F^{exA,emA} \ast \alpha) - (F^{exD,emD} \ast \beta) \]  \hspace{1cm} (6)

\[ \alpha = \frac{F^{exA,emA}}{F^{exA,emA}} \]  \hspace{1cm} (7)

\[ \beta = \frac{F^{exD,emD}}{F^{exD,emD}} \]  \hspace{1cm} (8)

- \( nF \): indicates the corrected fluorescence signal
- \( F \): indicates the fluorescence signal in presence of the donor and the acceptor
- \( F_A \): indicates the fluorescence signal in presence of the acceptor
- \( F_D \): indicates the fluorescence signal in presence of the donor
- \( exD \): donor excitation
- \( exA \): acceptor excitation
- \( emD \): donor emission
- \( emA \): acceptor emission

FRET efficiency can be described by the Förster equation (9). Förster radius \((R_0)\) describes the distance between donor and acceptor fluorophores, when 50% energy transfer takes place and it depends on the fluorophores used. The actual distance between donor and acceptor is given with \(r_0\). Formula (9) can be converted to calculate \(r_0\), if FRET efficiency is given (Formula 10).

\[ E = \frac{R_0^6}{R_0^6 + r_0^6} \]  \hspace{1cm} (9)

\[ r_0 = \sqrt[6]{\frac{R_0^6}{E - R_0^6}} \]  \hspace{1cm} (10)

A stoichiometric method which results in a FRET efficiency was described by Hoppe et al. [98]. They applied the correction factor \(\gamma\), which describes the relative excitability of the donor and acceptor at the donor excitation wavelength [244].

\[ E_f_A = \gamma \ast \frac{nF}{\alpha \ast F^{exA,emA}} \]  \hspace{1cm} (11)

\[ \gamma = \frac{\varepsilon_A^{exD}}{\varepsilon_D^{exD}} \]  \hspace{1cm} (12)
\( \varepsilon_A^{exD} \): extinction coefficient of the acceptor fluorophore at the donor excited wavelength

\( \varepsilon_B^{exD} \): extinction coefficient of the donor fluorophore at the donor excited wavelength

In the experiment we excited Cy3 at 520 nm to minimize cross talk. Thus, we needed the \( \varepsilon_A^{520\text{nm}} \) and \( \varepsilon_B^{520\text{nm}} \) to calculate the correction factor \( \gamma \).

Additional the donor quench (\( E_q \)) was expressed in percentage [%].

\[
E_q = \frac{F_{D}^{exD}emD - F_{D}^{exD}emD}{F_{D}^{exD}emD} \tag{13}
\]
Part III

RESULTS AND DISCUSSION
3.1 Characterization of Highly Specific Nucleases in *in vitro* and *in vivo*

This chapter characterizes the monomeric strand and site specific nickase AvrBs4-MutH under *in vitro* and *in vivo* conditions with a focus on the optimization of the target design. Additionally, AvrBs4-FokI and AvrBs3-PvuII variants were studied in *Saccharomyces cerevisiae* and compared to the golden standard meganuclease I-SceI. Furthermore, the RNA guided Cas9 protein was evaluated in yeast. Finally, the most promising TALEN with the optimized target design was applied to a therapeutic relevant gene causing X-linked retinitis pigmentosa.

3.1.1 Characterization of TALENs

3.1.1.1 AvrBs4-MutH

AvrBs4-MutH designed by Gabsalilow *et al.* [78], is a truncated AvrBs4 protein fused to the nickase MutH. The coding sequence of the N-terminus of AvrBs4 is shortened by 152 aa resulting in 138 aa and the C-terminus contains 28 aa. The nickase MutH, which addresses the 4 bp sequence (GATC), is fused to the C-terminus of AvrBs4 via two serine residues. The TALEN protein holds a Strep-tag at the N-terminus and a His-tag at the C-terminus, which enable a tandem purification (Figure 13).

![Architecture of AvrBs4-MutH](image-url)

Figure 13: Architecture of AvrBs4-MutH

AvrBs4 contains a N-terminus with 138 aa, a repeat domain with 595 aa, and a C-terminus with 28 aa. The Strep-tag is fused to the N-terminus, which consist of the non-canonical repeats N₀, N₁, and N₂. The binding domain encloses 17.5 repeats. Every repeat holds 34 highly conserved aa. In position 12 and 13 the aa residues are crucial for DNA recognition (bold). Two serines (S-S) connect the truncated AvrBs4 to the nickase MutH. The nickase carries a His-tag at the C-terminus and addresses a GATC site.
Previous *in vitro* experiments revealed that AvrBs4-MutH is capable to induce strand and site specific SSBs. Dependent on the spacer length between the AvrBs4 addressed site and the GATC sequence, the target is nicked on the bottom strand (3 bp spacer) or on the top-strand (6 bp spacer) (Figure 14). These characteristics of the nickase AvrBs4-MutH have been described by Gabsalilow *et al.* [78] in detail. For the present study, AvrBs4-MutH was analyzed on specifically designed target sequences. The construction of the new recognition sites were based on recent findings in the filed of genome targeting [53, 40, 199, 191]. In respect of HDR efficiency, the importance of nicking either the transcribed or the non-transcribed strand has been emphasized [53]. Accordingly, the target T-3-H, T-6-H and H-3-T were used to characterize AvrBs3-MutH under *in vitro* and *in vivo* conditions. Double nicking by specific nickases has been shown to reduce off-target activity [40, 199, 191]. This encouraged us to analyze the mechanism of double nicking by AvrBs4-MutH on the double targets T-3-HH-3-T and T-3-H-3-T. Three single targets (T-3-H; H-3-T and T-6-H) were designed to be nicked by AvrBs4-MutH and two double targets were created to detect double nicking. All targets comprise an I-SceI target site and are visualized in figure 14.

Figure 14: Design of specific single and double targets

I-SceI target site encloses 18 bp and can be found in all targets 17-21 bp in front of the binding site of AvrBs4-MutH. TALE binding site (T) is marked in green, whereas the recognition site of MutH (H) is shown in blue. The spacer length between the two binding sites is given in bp (x = 3, 6). Arrows demonstrate the nicking positions of MutH. T-3-H, H-3-T, and T-6-H are nicked (N) at a single position and are termed single targets. T-3-HH-3-T and T-3-H-T carry two positions to induce SSB and are termed double targets.
3.1.1.2 *in vitro* Analysis

The cutting mechanism and activity of AvrBs4-MutH were analyzed by employing a plasmid cleavage assay and a cleavage assay using different fluorescently 5'-end labeled PCR products. First, the inverted single target sites (T-3-H and H-3-T) were analyzed and compared with a plasmid cleavage assay (Figure 15). AvrBs4-MutH induced almost two times faster (k = 3.8 ± 0.6 x 10^-2 min^-1) SSBs to target H-3-T, than to target T-3-H (k = 2.2 ± 0.9 x 10^-3 min^-1).

![Figure 15: Kinetic cleavage assay on T-3-H and H-3-T](image)

Fraction of nicked DNA of T-3-H and H-3-T, analyzed by a plasmid cleavage assay. 8 nM substrate was incubated with a four fold excess of AvrBs4-MutH and time points were taken at 3', 10', 30', 60', 120', 180'. (A) Agarose gels (0.8%) stained with ethidium bromide. The gel demonstrates the super-coiled DNA, linear DNA, and open circle DNA. Substrates were digested by the nicking enzyme Nt.BspQ1 (N) and BamH1 (DSB), respectively. (B) Quantitative evaluation of the cleavage assay. Each measuring point represents the mean ± standard error indicated by error bars for each point. (n) number of independent experiments. AvrBs4-MutH had a cleavage rate constant for H-3-T of k = 3.8 x 10^-2 ± 0.6 x 10^-4 min^-1 and for T-3-H k = 2.2 x 10^-2 ± 0.9 x 10^-3 min^-1.
Next, we identified which DNA strand was nicked by AvrBs4-MutH using the cleavage assay on 5’-end labeled PCR fragments. The principle of the cleavage assay with fluorescently 5’-end labeled PCR products is depicted in figure 16 by using the example of the double target T-3-H-3-T. Relative to the strand which is cut on the fluorescently 5’-end labeled PCR product by AvrBs3-MutH, either a long fragment (e.g. 169 nt) with a red fluorescent signal or a short fragment (e.g. 59 nt) with a green fluorescent signal result.

First, the results of the single targets were analyzed on denaturing PAGE and native PAA gels. While the denaturing PAGE gel revealed which DNA strand is nicked, the occurrence of DSBs was detected by the analysis of native PAA gels. T-3-H was cleaved on the bottom strand, whereas H-3-T and T-6-H were nicked on the top strand (Figure 17 (A)). No DSBs were induced (Figure 17 (B)). The analysis of the fluorescently 5’-end labeled PCR double targets on a denaturing PAGE revealed that top and bottom strands were nicked in a 1:1 ratio (Table 19 and Figure 18 (A) and 19). No double strand breaks were detectable after 180 min of digestion. However, removing AvrBs4-MutH by Proteinase K (PK) and adding new AvrBs4-MutH after deactivation of PK led to double nicks in particular on T-3-H-3-T (Figure 18 (B)). The analysis of enzyme kinetics on double targets revealed that AvrBs3-MutH nicked T-3-H-3-T three to four times faster than T-3-HH-3-T (Figure 19) and (Table 19).
3.1 CHARACTERIZATION OF HIGHLY SPECIFIC NUCLEASES in vitro AND in vivo

Figure 17: In vitro characterization of AvrBs4-MutH on single targets (T-3-H, T-6-H and H-3-T)

Cleavage assay with fluorescently 5’-end labeled PCR products analyzed on a denaturing PAGE (A) and a native PAA gel (B). (-) uncut labeled PCR target (20 nM). (E) Target preincubated for 180 min with six fold excess of AvrBs4-MutH. (c) Target digested by BamHI. The uncut fragment is colored in orange. Fragments resulting from top strand nicking were labeled with a green fluorophore (Atto 488), while fragments resulting from bottom strand nicks were labeled with a red fluorophore (Atto 647N). T-3-H was nicked on the bottom strand, whereas T-6-H and H-3-T were nicked on the top strand. AvrBs4-MutH induced no DSBs on the single targets.

Figure 18: In vitro characterization of AvrBs4-MutH on double targets (T-3-H-3-T and T-3-HH-3-T).

Cleavage assay with 5’-end labelled PCR double targets (T-3-H-3-T and T-3-HH-3-T) analyzed on a denaturing PAGE (A) and a native PAA gel (B). (-) uncut labeled PCR target (20 nM); (E) target incubated with six fold excess of AvrBs4-MutH for 180 min; (P+E) AvrBs4-MutH preincubated on target site (E) was removed with PK (10 min, 37°C) and additional protein (120 nM) was incubated for 60 min after deactivation of PK; (c) control, target digested by BamHI (T-3-H-3-T) or ClaI (T-3-HH-3-T). The uncut fragment is shown in orange. Fragments resulting from top strand nicking were labeled with a green fluorophore (Atto 488), while fragments nicked at the bottom strand were labeled with a red fluorophore (Atto 647N). After treatment with PK and adding new AvrBs4-MutH after deactivation of PK DSBs were detectable in particular on T-3-H-3-T.
Figure 19: Kinetics of DNA nicking of AvrBs4-MutH on the double targets (T-3-H-3-T, T-3-HH-3-T)
Fluorescently 5'-end labeled PCR target (20 nM) was incubated with a six fold excess of AvrBs4-MutH and time points were taken at 3’, 10’, 30’, 60’, 120’, 180’. Each measuring point represents the mean ± standard error indicated by error bars for each point. (n) number of independent experiments. AvrBs4-MutH nicked T-3-H-3-T (A) three to four times faster than T-3-HH-3-T (B). Cleavage rate constants of AvrBs4-MutH (n=3): T-3-H-3-T: $0.023 \pm 2.5 \times 10^{-2}$ min$^{-1}$ (top strand), $0.029 \pm 2.3 \times 10^{-3}$ min$^{-1}$ (bottom strand); T-3-HH-3-T: $6.8 \times 10^{-3} \pm 9.4 \times 10^{-4}$ min$^{-1}$ (top strand), $6.9 \times 10^{-3} \pm 6.7 \times 10^{-4}$ min$^{-1}$ (bottom strand).
Table 19: Enzyme activity of AvrBs4-MutH for two independent double targets

Top and bottom strand-specific cleavage rate constant $K$ [min$^{-1}$] of AvrBs3-MutH is shown for the double targets T-3-HH-3-T and T-3-H-3-T. The binding site of AvrBs4 is marked with T, the GATC binding site of MutH is labeled with H, and the spacer length between the binding sites is given in bp.

<table>
<thead>
<tr>
<th>Double Target</th>
<th>Strand</th>
<th>$K$ [min$^{-1}$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top strand</td>
<td>T-3-HH-3-T</td>
<td>6.8 x 10$^{-3}$ ± 9.4 x 10$^{-4}$</td>
</tr>
<tr>
<td>Bottom strand</td>
<td>T-3-H-3-T</td>
<td>6.9 x 10$^{-3}$ ± 6.7 x 10$^{-4}$</td>
</tr>
</tbody>
</table>

3.1.1.3 in vivo Analysis

To challenge the established specific nucleases (I-SceI or TALE-FokI), TALE-MutH has to provide excellent in vivo efficiency. To investigate the efficiency of TALE-MutH under in vivo conditions and to validate the in vitro results, a single strand in vivo assay in yeast (Figure 20), as has been previously reported by Cermak et al. [33], was employed. S. cerevisiae is an eukaryotic microorganisms, providing a fast and cost efficient model for in vivo conditions by using the single strand annealing assay. The successful cleavage of the target positioned between two partially duplicated fragments of the lacZ gene reconstituted a functional lacZ gene is reconstituted by subsequent single strand annealing (SSA) (Figure 11). This experimental approach enables a quantitative analysis of the nuclease activity (Figure 20). The functional lacZ gene encodes a lacZ fragment which dimerize with the lacZ deletion mutant resulting in an active β-gal.

The β-gal activity of TALE-MutH was compared to I-SceI and AvrBs4-FokI. I-SceI induces DSBs within its non-palidromic 18 bp target sequence and is the established GOLDEN STANDARD due to its high specificity combined with high nuclease activity. The homodimeric AvrBs4-FokI induces DSBs within a 36 bp (2 x 18 bp) substrate. The repeat array of AvrBs4 was introduced to a truncated variant of the TALE protein via the GOLDEN GATE cloning strategy (Methods 2.2.1.1). The AvrBs4-MutH design was similar to the in vitro design, but lacking the His- and Strep-tag. Similar to AvrBs4-MutH, the N-terminus of AvrBs4-FokI lacks the coding sequence of 152 aa residues, but contains a 68 aa as linker to the unspecific cleavage domain FokI. A yeast expression plasmid with AvrBs4 without catalytic domain served as negative control.

Additionally, several variants of AvrBs3-PvuII were analyzed by the SSA assay, as PvuII proofed to be a promising alternative to the unspecific cleavage domain FokI. The specific nuclease PvuII increases the number of bp in the target site (42 bp) [239].
Figure 20: Single strand annealing assay
A yeast based recombinant assay was used to analyze the efficiency of the specific nucleases. The specific nucleases bind to the recognition sequence, which is flanked by two partially duplicated fragments of the lacZ gene (blue), SSA is triggered and a functional lacZ gene is reconstructed. Right corner: Yeast colonies (blue) with I-SceI expression plasmid and target plasmid growing on a matrix containing X-Gal.

3.1.1.4 AvrBs4-MutH

Two versions of the viability assay were used. First, yeast cells were transformed with expression plasmids, but no target plasmid was present. The total amount of yeast colonies of AvrBs4 without nuclease were compared to I-SceI, AvrBs4-FokI and AvrBs4-MutH and no significant difference in the total amount of colonies could be detected (Figure 21 (A)). Second, yeast cells were transformed with expression plasmids in the presence of the plasmid with the defined target for the specific nucleases. This resulted in a significant reduction of the total amount of yeast colonies compared to the total amount of yeast colonies transfected with AvrBs4 without nuclease activity, except in the case of AvrBs4-MutH on the target T-3-H and H-3-T (Figure 21 (B)). AvrBs4-FokI revealed the lowest number of colonies, followed by AvrBs4-MutH on the double target T-3-HH-3-T. However, when compared to I-SceI, AvrBs4-MutH with the targets T-3-H-3-T, T-3-H, and H-3-T led to higher colony counts, while colony counts with the target T-6-H were comparable to I-SceI.
3.1 Characterization of Highly Specific Nucleases in vitro and in vivo

Figure 21: Relative survival of yeast colonies

In vivo characterization of AvrBs4-MutH as nicking and double nicking enzyme compared to established specific nucleases (I-SceI, AvrBs4-FokI). The specific nucleases are compared to AvrBs4 without nuclease activity and significant differences were determined by the unpaired T-test (n ≥ 3, n.s. = not significant, P < 0.05 *, P < 0.01 **, P < 0.001 ***). Targets underlined with DN (double nick) carry two positions to induce a SSB, while N (nick) indicate a single position to induce SSB.

(A) Total amount of yeast colonies transfected with expression plasmid for the specific nuclease without the target plasmid.

(B) Quantification of yeast colonies transfected with the specific nuclease and the specific target plasmid.
To analyze the specific activity of the nucleases on different targets, the ratio between blue and unstained colonies was calculated and compared to AvrBs4 without nuclease activity (Figure 22 (A and B)).

Figure 22: β-gal activity of specific nucleases (I-SceI, AvrBs4-FokI, and AvrBs4-MutH) in *S. cerevisiae*. In *vivo* characterization of AvrBs4-MutH as nicking and double nicking enzyme compared to established specific nucleases (I-SceI, AvrBs4-FokI). The specific nucleases are compared to AvrBs4 without nuclease activity and significant differences were determined by the unpaired T-test (*n* ≥ 3, n.s. = not significant, *P* < 0.05 *, *P* < 0.01 **, *P* < 0.001 ***). Targets underlined with DN (double nick) carry two positions to induce a SSB, while N (nick) indicate a single position to induce SSB. (A) Fraction of blue colonies expressed as percentage [%] of the different specific nucleases. (B) Relative number of white and blue colonies of the nucleases on their specific targets.
When compared to AvrBs4 all constructs on all target combinations showed a significantly higher β-gal activity. Even the single targets showed a significantly higher β-gal activity. This finding indicates that nicking can trigger SSA. The golden standard I-SceI and AvrBs4-MutH addressing the double target T-3-H-3-T resulted the highest activity, which approximated 100% of stained colonies (I-SceI: 100 ± 1%; T-3-H-3-T: 99 ± 1%), followed by AvrBs4-MutH on the double target T-3-HH-3-T (T-3-HH-3-T: 95 ± 2%) and AvrBs4-FokI (AvrBs4-FokI: 83 ± 17%). The single target T-6-H (8 ± 3%); revealed a three to four fold lower activity when compared to T-3-H (84 ± 9%) and H-3-T (84 ± 2%) (Figure 22 (A)).

Next, the number of white and blue colonies relative to the results of AvrBs4 were analyzed (Figure 22 (B)). Here, AvrBs4-FokI and AvrBs4-MutH targeting T-6-H showed relatively low activities, while the highest enzyme activities were detected for the combinations of AvrBs4-MutH with T-3-H-3-T, T-3-H, and H-3-T. This indicates that AvrBs4-MutH has a greater in vivo efficiency compared to the golden standard I-SceI. While incubation over two additional days led to a further increase in the intensity of the blue-staining of the T-3-H and H-3-T colonies, all other constructs did not change the intensity of the blue color (Figure 23).

Figure 23: Yeast colonies growing on X-gal containing matrix
Clipping from the image of X-GAL containing plates. Yeast cells (YPH500) were co-transformed with a 1:1 ratio of expression plasmid and reporter plasmid. Cells were allowed to grow for 5 days (30°C) on media plates, containing 0.08 mg/ml X-GAL. AvrBs4-MutH on T-3-H (left) and AvrBs4-MutH on H-3-T (center) resulted in white, light blue, and dark blue colonies. As example, a light blue colony and a dark blue colony are marked in red (left). I-SceI on T-3-HH-3-T containing the I-SceI recognition sequence (right) let to dark blue colonies.
3.1.1.5 *AvrBs3-PvuII* variants

It has been recently shown that the catalytic unspecific domain of FokI can also be replaced by the DNA specific cleavage domain PvuII [239]. To investigate the efficiency of the “tool box” of TALENs, AvrBs3-PvuII variants (Figure 24) were analyzed using the yeast based recombinant single strand annealing assay.

![Figure 24: AvrBs3-PvuII variants](image)

The binding domain AvrBs3 is fused to the specific cleavage domain PvuII via a linker containing the remaining 28 aa of the C-terminus from AvrBs3, a 14 aa linker from the FokI protein and two additional serines. The three TALE-PvuII variants analyzed by the SSA assay in *Saccharomyces cerevisiae* are shown. (I) AvrBs3-hf-PvuII, a homodimeric TALEN, containing the high fidelity (hf) variant of PvuII(T46G) and addressing a tripartial substrate of 42 bp. (II) AvrBs3-sc-PvuII holds homodimeric PvuII domains, which are fused to each other by a 4 aa linker (GSGG). The single chain variant target a bipartial sequence with 24 bp and contains the T46G mutation. (III) AvrBs3-hd-PvuII, a heterodimeric TALEN, includes the T46G mutation and either a PvuII domain with the substitution D30K or a converse substitution K38D. Together, the two PvuII variants yield an active PvuII nuclease.

AvrBs3-PvuII coding sequences were imported into the yeast expression plasmid. Similar to the previous TALEN constructs, the AvrBs3 repeat array (Figure 33 (A)) was established by employing the GOLDEN GATE cloning strategy (Methods 2.2.1.1) and introduced into a truncated TALE protein (NΔ152-C28). Additional to the C28 linker from the TALE protein, a 14 aa linker based on the FokI-linker (VIP-NRGVTKQLVKG) [225] and two additional serines link the binding module to the cleavage domain. This design was previously used for ZFNs [190] and adopted by Yanik et al. [239] for TALE-PvuII variants. Three distinct AvrBs3-PvuII variants were characterized by the single strand annealing assay and are described in Figure 24. (I) AvrBs3-hf-PvuII [239] is a homodimeric TALEN containing the high fidelity (hf) variant of PvuII(T46G) to prevent star site activity and it addresses a tripartial substrate. Based on this architecture, the single chain vari-
Figure 25: Recognition sequence of AvrB3-PvuII variants

All AvrBs3-PvuII variants were analyzed on the T-4-P-4-T target. Including the 6 bp target site of PvuII (P) flanked by the target sequence of AvrBs3 (T). The spacer length between the two target sequences is given in bp and the arrow refer to the cutting position of PvuII.

Variant (II) AvrBs3-sc-PvuII [239] was designed and includes a T46G substitution. The homodimeric PvuII domains of AvrBs3-sc-PvuII (II) are linked to each other by a 4 aa linker (GSGG) [202] and the monomeric TALEN recognizes a bipartial sequence. (III) AvrBs3-hd-PvuII [238] is a heterodimeric (hd) TALEN carrying a PvuII domain, including a D30K substitution or a PvuII domain, which contains a K38D mutation. Together, the two PvuII domains built an active PvuII cleavage domain. All AvrBs3-variants induce DSBs and were analyzed on the tripartial target T-4-P-4-T (Figure 25), which has been reported to be the optimal substrate design [239].

Beginning with the analysis of viability, equal amounts of expression vector were introduced to the yeast cells (Figure 26 (A)). The I-SceI results were set at 1.0 and all further results were correlated to I-SceI. The total count of yeast colonies of the specific AvrBs3-PvuII variants were compared to the count of yeast colonies of AvrBs4 without nuclease activity. AvrBs3-hf-PvuII (I) and AvrBs3-sc-PvuII (II) resulted in a significant reduction (P < 0.001 ***) of the number of yeast colonies, when compared to the colony numbers of AvrBs4. The count of yeast colonies of AvrBs3-hd-PvuII (III) compared to AvrBs4 showed no significant differences. Furthermore, no significant different number of yeast colonies were detected, when single substitutions of the heterodimeric TALEN (e.g. AvrBs3-K38D-PvuII) were analyzed individually and compared to AvrBs4.

Next, the expression vector, as well as the target plasmid were introduced into the yeast cells and the amount of yeast colonies were analyzed (Figure 26 (B)). Again AvrBs3-hf-PvuII (I) and AvrBs3-sc-PvuII (II) yielded a significant reduction (P < 0.05 *) of the number of yeast colonies compared to AvrBs4 without nuclease activity. The total amount of yeast colonies of AvrBs3-hd-PvuII and AvrBs3-K38D-PvuII showed a significant decrease (P < 0.05 *) when compared to the values of AvrBs4. AvrBs3-D30K-PvuII had no significant different number of yeast colonies, when compared to AvrBs4.
Figure 26: Relative survival of yeast colonies

The number of yeast colonies relative to the results of I-SceI is given. The specific nucleases are compared to AvrBs4 without nuclease activity and significant differences were determined by the unpaired T-test (n ≥ 3, n.s. = not significant, P < 0.05 *, P < 0.01 **, P < 0.001 ***). (A) Relative amount of yeast colonies transformed with expression plasmid for the specific nuclease without any target. (B) Quantification of yeast colonies transfected with the specific nuclease and the specific target plasmid.

Next, the β-gal activity of the specific nucleases was determined. First, the ratio between white (no functional β-gal ) and blue (functional β-gal ) colonies was calculated and compared to AvrBs4 without nuclease activity (Figure 27). AvrBs3-hf-PvuII (I) and AvrBs3-sc-PvuII (II) revealed no significant different β-gal activity, when compared to AvrBs4 (percentage of blue colonies: 4 ± 2%), while AvrBs3-
hd-PvuII (III) had a significantly higher \( \beta \)-gal activity (\( P < 0.001^{***} \); percentage of blue colonies: \( 63 \pm 10\% \)). Still, AvrBs3-K38D-PvuII reached a higher \( \beta \)-gal activity (\( 95 \pm 3\% \)) when compared to the heterodimeric construct. AvrBs3-D30K-PvuII uncovered no significantly higher \( \beta \)-gal activity (\( 2 \pm 1\% \)) compared to the AvrBs4 results. The relative number of white and blue colonies is shown in figure 28. AvrBs3-hf-PvuII (I) and AvrBs3-sc-PvuII (II) revealed no significant differences in enzyme activity when compared to AvrBs4 without nuclease (relative number of blue colonies: \( 0.03 \pm 0.01 \)). AvrBs3-hd-PvuII (III) (relative number of blue colonies: \( 0.2 \pm 0.1 \)) and the individual analysis of AvrBs3-K38D-PvuII (relative number of blue colonies: \( 0.3 \pm 0.2 \)) disclosed a significantly higher number of blue colonies (\( P < 0.01^{**} \)) compared to the relative number of blue colonies of AvrBs4. The individual analysis of AvrBs3-D30K-PvuII (relative number of blue colonies: \( 0.01 \pm 0.01 \)) showed no significantly increased activity compared to AvrBs4 without nuclease activity.

Figure 27: Relative \( \beta \)-gal activity of specific nucleases (I-SceI and AvrBs3-PvuII variants) in \textit{S. cerevisiae} as analyzed by the ratio of blue colonies expressed in percentage [%]

Single Strand Annealing assay in yeast. All results were normalized to I-SceI. Levels of significance were determined by unpaired T-tests (\( n \geq 3 \), n.s. = not significant, \( P < 0.05^{*} \), \( P < 0.01^{**} \), \( P < 0.001^{***} \)). Percentage of blue colonies [%] of the different specific nucleases. Relative number of white and blue colonies of the nucleases on their specific targets.
Figure 28: β-gal activity of specific nucleases (I-SceI and AvrBs3-PvuII variants) in *S. cerevisiae*

The specific nucleases are compared to AvrBs4 without nuclease activity and significant differences were determined by the unpaired T-test ($n \geq 3$, n.s. = not significant, $P < 0.05$ *, $P < 0.01$ **, $P < 0.001$ ***). Relative number of white and blue colonies of the nucleases on their specific targets.

3.1.2  **Characterization of RNA guided Cas9 protein in vivo**

Due to simple handling and the ability of multiplexing RNA guided Cas9 nuclease appears to be a promising new specific nuclease [127, 191, 249]. The β-gal activity of a type II bacterial Cas9 protein on a target overlapping the I-SceI recognition sequence (Figure 29), was analyzed by employing the SSA assay in *S. cerevisiae* (Figure 30).

The spCas9 kindly supplied by George Church [58] addresses a 20 bp sequence overlapping the I-SceI target in the T-3-H-3-T substrate. At the end of the 20 bp target sequence, the required 5’-NGG protospacer-adjacent motif (PAM) is placed (Figure 30). The architecture of the gRNA is based on the empty *S. cerevisiae* expression vector from Timothy Lu [69]. The 20 nucleotides (nt) complementary to the DNA target were introduced to the empty expression vector by using the Gibson assembly cloning strategy. Analyzing the Cas9::gRNA system in yeast revealed a significantly higher β-gal activity ($P < 0.001$ ***; β-gal activity: 23 ± 0.1%) compared to AvrBs4 without nuclease activity (Figure 30).
3.1 Characterization of Highly Specific Nucleases in vitro and in vivo

Figure 29: Architecture of Cas9::gRNA and its recognition sequence
(A) Cas9::gRNA addresses a 20 bp target which overlap with the I-SceI target sequence in T-3-H-3-T. At the end of the 20 bp target the 5’-NGG protospacer-adjacent motif (PAM) is placed. gRNA, carrying 20 nt complementary to the DNA target sequence recognizes its target via Watson-Crick base pairing. (B) Target of I-SceI is illustrated in bold letters. The addressed sequence of Cas9::gRNA is underlined in black and the PAM sequence is displayed in orange. The addressed site is positioned 9 bp in front of the target site (T-3-H-3-T) of AvrBs4-MutH.

Figure 30: ß-gal activity of Cas9::gRNA in S. cerevisiae
On the left site the ß-gal activity of Cas9::gRNA is compared to AvrBs4 without nuclease activity and significant differences were determined by the unpaired T-test (n ≥ 3, n.s. = not significant, P < 0.05 *, P < 0.01 **, P < 0.001 ***). On the right site the absolute counts of white and blue colonies are shown.
3.1.3 Alignment search of the addressed sites in the genome of *Saccharomyces cerevisiae*

One key limitation of the specific nucleases is off-target activity. We therefore searched the yeast genome for recognition sites analyzed in the SSA assay. Alignments in the genome of *S. cerevisiae* to the addressed sequences were search by using the Basic Local Alignment Search Tool (BLAST) [5]. The S288c strain is the most widely applied yeast strain and its genome has been completely sequenced. YPH500 is driven from the YNN216 strain, which is congenic to the S288c strain and differs only in one locus and a linked segment of chromosome [200].

First, no hits for any of the recognition sequences were found within the genome of *S. cerevisiae* holding an identity of 100%. It has been described that specific nucleases can address a target site including mismatches. How many mismatches are tolerated depends on the family of the specific nuclease and the location of the mismatch within the target sequence. Beginning with the I-SceI recognition site, which consist of 18 bp, the best hit was found in the chromosome XII (National Center for Biothechnology Information (NCBI) reference sequence: NC_001144.5) with an identity of 77% (14 nt / 18 nt), but missing the 3'-end. It is known, that I-SceI has a two step recognition mechanisms. First, a strong binding take place to the 3'-end of the target (seed region) followed by binding of I-SceI to the 5'-end region [168]. Consequently, mismatches at the 3'-end have a high impact on the specificity of I-SceI. Thus, alignments were searched which included at least the 9 bp seed region at the 3'-end. The two best alignments, including the 3'-end, were identified in the chromosome XVI (NCBI reference sequence: NC_001148.4) (identity: 66% (12 nt / 18 nt); Table 20). Furthermore, single hits were found with an identity to the I-SceI target sequence of 61% (11 nt / 18 nt) in chromosomes XII (NCBI reference sequence: NC_001144.5) and XIII (NCBI reference sequence: NC_001145.3).

Next, the target sequence of AvrBs4-MutH (T-3-H) was analyzed. In contrast to I-SceI, TALE proteins bind more efficiently to the 5'-end of the target [142]. There were several hits to the AvrBs4 target with an identity of 68% (13 nt / 19 nt), including the 10 bp 5'-end of the AvrBs4 recognition site (Table 20). But there was no AvrBs4 recognition site next to a GATC site with a distance of 3 bp. Furthermore, there was no sequence containing the GATC site and the 5'-end of the TALE recognition site.
Table 20: Identity of the best hits enclosing the seed region of the specific target site

The size of the specific target site of each investigated specific nuclease are given in bp. The seed region can be located at the 5’end or at the 3’end of the recognition site. TALE binding site in combination with the additional specific sequence of the nuclease do not exist (n.e.: non-existence).

<table>
<thead>
<tr>
<th>specific nuclease</th>
<th>target size [bp]</th>
<th>5’end seed region</th>
<th>3’end seed region</th>
<th>+ nuclease target</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-SceI</td>
<td>18 bp</td>
<td>-</td>
<td>66</td>
<td>-</td>
</tr>
<tr>
<td>AvrBs4-MutH</td>
<td>18 bp</td>
<td>68</td>
<td>-</td>
<td>n.e.</td>
</tr>
<tr>
<td>AvrBs3-PvuII</td>
<td>18 bp</td>
<td>68</td>
<td>-</td>
<td>n.e.</td>
</tr>
<tr>
<td>Cas9::gRNA</td>
<td>20 bp + 5’ NGG</td>
<td>-</td>
<td>52</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>20 bp + 5’NAG</td>
<td>-</td>
<td>52</td>
<td>-</td>
</tr>
</tbody>
</table>

Furthermore, a fraction of the target site of AvrBs3-PvuII (T-4-P) was analyzed. One alignment was detected in chromosome XII, including the 10 bp 5’-end containing an identity for the AvrBs3 sequence of 68% (13 nt / 19 nt), but not containing the PvuII-site (Table 20). The best hit to the AvrBs3 target site was found in chromosome VII with an identity of 79% (15 nt / 19 nt), yet missing the 5’-end sequence and the PvuII recognition site.

Finally, the sequence addressed by Cas9::gRNA was analyzed. To target a sequence with the Cas9::gRNA construct a 5’-NGG a 5’-NAG PAM is required [107]. Additionally, Cas9::gRNA is more sensitive to mismatches at the 3’-end (15 bp to the PAM) [108, 47, 107]. The yeast genome was searched for hits including one of the PAMs and parts of the 3’-end target sequence. Three alignments containing the 5’-NGG PAM motif were detected, one in chromosome VIII (NCBI reference sequence: NC_001140.6) with an identity of 52% (12 nt / 23 nt) (Table 20) and two with an identity of 48% (11 nt / 23 nt) in distinct chromosomes. Similar results were found with a 5’-NAG PAM (Table 20). One hit was detected in the chromosome IX (NCBI reference sequence: NC_001141.2) with an identity of 52% (12 nt / 23 nt) and two alignments with an identity of 48% (11 nt / 23 nt) were found in chromosome II (NCBI reference sequence: NC_001143.8) and in chromosome XI (NCBI reference sequence: NC_001143.9).
3.1.4 Addressing X-linked Retinitis Pigmentosa with specific nucleases

3.1.4.1 Design of TALE-MutH proteins addressing the retinitis pigmentosa GTPase regulator (RPGR) gene

Due to the promising results of the SSA assay, TALE-MutH was re-designed to target the retinitis pigmentosa GTPase regulator (RPGR) gene, which is involved in the inherited eye disease X-linked retinitis pigmentosa (RP) (Table 17). In figure 31 the constructed TALENs are illustrated.

Figure 31: Sequence within the RPGR gene targeted by the designed TALE-MutH variants

The illustrated sequence is a part of ORF15 within the RPGR gene. The GATC site (I) (blue) is addressed by three of the five constructed TALE-MutH variants (TALE(RPGR)-1 to 3 (I)). Two of them (TALE(RPGR)-1(I) and TALE(RPGR)-2(I) address a target site which contain a 3 bp spacer between the TALE recognition site and the GATC site (I). TALE(RPGR)-3(I) recognizes a sequence including a 6 bp spacer between the two recognition sites. For the second GATC site (II) (light blue) two additional TALE-MutH variants were created. Recognition site of TALE(RPGR)-1(II) includes a 3 bp spacer and is therefore nicked on the sense strand. Due to its 6 bp spacer TALE(RPGR)-2(II) nicks the anti-sense strand on the target sequence.

Similar to the naturally occurring TALE binding domains (AvrBs4 and AvrBs3), the new binding domains were reconstructed with the *golden gate* (methods 2.2.1.1) cloning strategy and introduced into an eukaryotic expression vector containing the coding sequence of MutH. Mutations causing X-linked retinitis pigmentosa are located inside the ORF15 of the RPGR gene. The TALENs were designed to address a sequence within the ORF15 nearby a GATC site. A correct DNA matrix will be introduced to the gene for gene editing. This can occur by inducing DSBs, SSBs, or by double nicking. Three TALE-MutH were designed to address the GATC site (I). TALE(RPGR)-1(I)-MutH and TALE(RPGR)-2(I)-MutH contain a 3 bp spacer on their addressed sites in between the recognition site of TALE and the GATC site. The design is based on the target site construction T-3-H-3-T. TALE(RPGR)-3(I)-MutH also addresses the GATC site (I), but recognizes the target site with a 6 bp spacer. This result in two heterodimeric TALE(RPGR)-MutH (TALE(RPGR)-1(I)-MutH / TALE(RPGR)-2(I)-MutH and TALE(RPGR)-2(I)-MutH / TALE(RPGR)-3(I)-MutH), which are able to induce double nicking within the ORF15. The second GATC site (II) can be addressed by
TALE(RPGR)-1(II)-MutH and TALE(RPGR)-2(II)-MutH. TALE(RPGR)-1(II)-MutH addresses the GATC site (II) with a 3 bp spacer in between the TALE target site and GATC site, whereas TALE(RPGR)-2(II)-MutH contains a spacer distance of 6 bp in between the TALE target site and GATC site. The heterodimeric TALE-MutH introduces a nick into the bottom strand and into the upper strand. Additionally, both GATC sites (I and II) can be addressed (e.g. TALE(RPGR)-1(I)-MutH / TALE(RPGR)-1(II)-MutH) resulting in a double nick, 13 bp apart from each other. Finally, all TALE-MutH constructs can also be applied individually to induce one SSB into the ORF. Two more GATC sites are located in the ORF, which represent further possible target for TALE-MutH constructs.

3.1.4.2 Characterization of TALE-MutH specific nucleases for the RP

3.1.4.3 GTPase regulator gene

The heterodimer TALE(RPGR)-1 (I) -MutH / TALE(RPGR)-2 (I) -MutH described in section 3.1.4.1 were analyzed under in vivo conditions in cooperation with the group Retinal gene therapy of Prof. Dr. Dr. Knut Stieger at the University of Giessen (FB 11). The specific nucleases were analyzed together with Dr. Mert Yanik episomal in HEK293T cells. Due to a DNA break NHEJ is triggered and a frame shift occurs, which results in a functional gene expressing blue fluorescent protein (BFP). The BFP was detected by flow cytometry and the results were compared to the GOLDEN STANDARD I-SceI (Figure 32 (A)). Analysis of TALE(RPGR)-1 (I) and TALE(RPGR)-2 (I) addressing the GATC (I) site revealed a similar NHEJ activity, when compared to I-SceI.

Additionally, cell toxicity studies were performed, by co-transfection of mCherry and the specific nucleases expression plasmids (Figure 32 (B)). A pair of ZF-FokI [6] and the TALE-MutH pair (TALE(RPGR)-1 (I)/TALE(RPGR)-2 (I)) were analyzed using the cell toxicity assay.

The empty control vector (pcDNA3) served as negative control. The ZF-FokI (survival rate: 42 ± 18%) pair led to a significantly reduced survival rate when compared to I-SceI (survival rate: 100 ± 8%). Whereas the heterodimeric TALE(RPGR)-MutH (survival rate: 89 ± 15%) resulted in a comparable survival rate to I-SceI. The empty vector pcDNA3 (survival rate: 100 ± 8%) showed a similar survival rate as I-SceI.
**Figures 32:** Analysis of TALE-MuH addressing GATC (I) site within the ORF15 of the RPGR gene

Levels of significance were determined by unpaired T-tests (n ≥ 3, n.s. = not significant, P < 0.05 *, P < 0.01 **, P < 0.001 ***). (A) NHEJ activity analysis of TALE(RPGR)-1 (I) and TALE(RPGR)-2 (I) addressing the GATC (I) site, compared to NHEJ activity of I-SceI. NHEJ activity results in a functional gene, expressing BFP. (B) Cell toxicity was analyzed by the co-transfection of mCherry and specific nucleases expression plasmids. Survival rate was defined as the decrease of mCherry-positive cells from day 2 to day 5. Number of mCherry-positive cells were detected by flow cytometry and data were normalized to the I-SceI results. The empty vector (pcDNA3) served as negative control.
3.2 CHARACTERIZATION OF BINDING CHARACTERISTICS OF AVRBS3

This chapter provides a detailed analysis of the TALE binding protein AvrBs3(Δ N152-C28). The AvrBs3(Δ N152-C28) protein was re-designed to enable labeling at specific positions in the protein, thus resulting in several AvrBs3 variants. These new designed TALEs provide promising possibilities to analyze the DNA binding properties of TALEs and facilitate the analysis of the off-target problematic.

3.2.1 Variants of AvrBs3

The modified protein AvrBs3(Δ N152-C28) is a truncated version of the naturally occurring AvrBs3. The N-terminus lacks 152 aa and the C-terminus holds 28 aa. All AvrBs3(Δ N152-C28) variants carry a strep-tag at the N-terminus and were generated by the GOLDEN GATE cloning strategy (methods 2.2.1.1). Based on the mechanism of a thiol Michael-type addition reaction cysteine residues can react with the double bond of maleimide to succinimide-thiolethers. This mechanism can be used to label cysteine residues with maleimide-fluorophores. Thus, various AvrBs3 mutants were created and labeled at distinct positions (Figure 33).

![Architecture of AvrBs3 variants](image)

Figure 33: Architecture of AvrBs3 variants

All AvrBs3 variants contain a N-terminus with 138 aa, a repeat domain with 598 aa and a C-terminus with 28 aa. The Strep-tag is fused to the N-terminus, which comprises N₀, N₁, and N₂. The binding domain encloses 17.5 repeats. Each repeat holds 34 highly conserved aa. In position 12 and 13 the aa are crucial for DNA recognition (bold). Each repeat contains one cysteine residue located at position 30. In the newly constructed variants, cysteine residues are mutated to methionine in each repeat except for repeat number 1 and/or 16 (blue) depending on the specific AvrBs3 variants (AvrBs3 (A), AvrBs3 (B), and AvrBs3 (AB)).

The AvrBs3 binding domain contains 17.5 repeats and each repeat consists of 34 highly conserved aa, which enclose a cysteine residue in position 30 (Figure 34 (A)). To permit labeling at defined positions, the cysteine residues in the specific repeats were substituted to methionine residues via site-directed mutagenesis PCR.
The web server PoPMuSiC [55] was used to estimate the stability change of these single-site mutations of cysteine. It revealed that methionine has the smallest shift in its thermodynamic stability ($\Delta \Delta G = 0.3$ [kcal/mol]) when compared to cysteine (Figure 34 (B)). While methionine belongs to the non-polar aa and cysteine to the polar aa, they both include an alkyl chain and a sulfur atom.

![Figure 34: Single-site mutation of the cysteine residue](image)

(A) Crystal structure of AvrBs3 bound to DNA with a focus to one repeat is shown (pdb: 2YPF) [205]. The alpha-turn-alpha structure of this repeats is exhibited and aa 12 (H) and 13 (D) (orange) recognize the nucleotide cytosine. The cysteine residue at position 30 in each repeat is marked in blue and its thiol group in orange. (B) Prediction of AvrBs3 mutant stability changes $\Delta \Delta G$ [kcal/mol]. Methionine has the lowest stability change with $\Delta \Delta G = 0.3$ [kcal/mol].

Having the choice between plasmids encoding each repeat either with or without cysteine residues allowed to generate AvrBs3 variants with different numbers and positions of cysteine residues in its binding domain. Beside the variant AvrBs3($\triangle$ N152-C28), which contains a cysteine residue in each repeat, three additional AvrBs3($\triangle$ N152-C28) variants were created (Figure 33). Two of these AvrBs3($\triangle$ N152-C28) variants ((A) and (B)) contain a single cysteine residue in the entire protein and all remaining cysteine residues were mutated to methionine. AvrBs3 (A) encloses a single cysteine residue in its first repeat and AvrBs3 (B) has its single cysteine residue in repeat sixteen. These two variants enable to label AvrBs3($\triangle$ N152-C28) with a single fluorophore at the beginning of the binding domain (AvrBs3 (A)) or at the end of the binding domain (AvrBs3 (B)). Furthermore, an AvrBs3($\triangle$ N152-C28) variant was generated which allows labeling simultaneously at two positions in the binding domain (AvrBs3 (AB)). This variant carries two cysteine residues, one cysteine is located in the first repeat and one cysteine in repeat sixteen.
To enable labeling, the cysteine residue has to be accessible for the used dye in a way that it does not influence the protein functionality. Regarding the crystal structure of AvrBs3 the cysteine residue at position 30 in each repeat is in an exposed position and should be available for labeling. Figure 34 (A) shows the crystal structure of one repeat unit and the location of the cysteine residue is marked.

Purification of the AvrBs3(Δ N152-C28) variants was examined by affinity purification using a strep-tag fused to their N-terminus. In figure 35 (A), an example of a strep-tag purification of AvrBs3 (B) analyzed with a 12% SDS-polyacrylamide gel electrophoresis (SDS-PAGE) is given. AvrBs3 protein has a molecular weight of 80 kDa and was analyzed on the 12% SDS-polyacrylamid gel above 85 kDa. TALEs used for characterization of unlabeled AvrBs3 variants (e.g. determination of $K_D$-values) were purified in an additional step, using heparin sepharose affinity chromatography. The AvrBs3(Δ N152-C28) variants eluded at a NaCl concentration of 285 nM (Figure 35 (B)).

Figure 35: Purification of AvrBs3 (B)
(A) Strep-tag purification, taking the example of AvrBs3. This 12% SDS-polyacrylamide gel shows the marker (PageRuler Unstained Protein Ladder) and each working step of the strep-tag purification: Marker (M), flow through (FT), washing step 1 and 2 (W1 and W2), elution steps (E1-5), and dialysis (D). AvrBs3 band appears above the 85 kDa band. (B) Heparin sepharose affinity chromatography: In a second step AvrBs3 was purified over a heparin sepharose column. A conductivity gradient with high salt buffer B (0-100%) was created and AvrBs3 eluted at a NaCl concentration of 285 nM. Blue = absorption [mAU] at 280 nm; magenta = conductivity [%], red = fractions.
3.2.2 Binding character of the AvrBs3 variants

The K\text{D}-values of AvrBs3(\Delta N152-C28) and AvrBs3 (\(B\)) were determined by fluorescence anisotropy (Figure 36 and Figure 37). Protein was added in increasing concentrations to 5’-end Hex labeled specific DNA (0.5 nM). The K\text{D}-values of both AvrBs3 proteins were about 1 nM (AvrBs3(\Delta N152-C28): 1.9 ± 0.8 nM , AvrBs3 (B): 0.9 ± 0.3 nM (Table 21; Figure 36).

Figure 36: K\text{D}-values of AvrBs3(\Delta N152-C28) and AvrBs3 (B)

K\text{D}-values were analyzed by fluorescence anisotropy. A 25 bp long specific DNA (0.5 nM) labeled with HEX at the 5’-end of the bottom strand was used. Labeling position of HEX is marked in yellow in the crystal structure of AvrBs3 bound to specific DNA (pdb: 2YPF) [205]. TALE variants were titrated with increasing concentration to specific DNA. Each measuring point represents the mean ± standard error indicated by error bars for each point. n : number of independent experiments. Measuring points were fitted with a non-linear regression. K\text{D}-values determined: 1.9 ± 0.8 nM (AvrBs3(\Delta N152-C28)) and 0.9 ± 0.3 nM (AvrBs3 (B)).

The active fraction of protein was determined by titrating the protein with a total DNA concentration (10 nM) much higher than the K\text{D}-values (Figure 37). Under these conditions the K\text{D}-value cannot be determined, but the number of binding positions can be analyzed. The experiment revealed, that 13% of AvrBs3(\Delta N152-C28) and 15% of AvrBs3 (B) of the purified stock solutions were active. The calculated K\text{D}-values of 100% active protein were 0.23 ± 0.07 nM for AvrBs3(\Delta N152-C28) and 0.14 ± 0.05 nM for AvrBs3 (B) (Table 21).

Next, the rate constants for dissociation were analyzed for the AvrBs3 variants. Here, the 5’-end HEX labeled specific DNA competed with
Figure 37: Determination of active fractions of AvrBs3(ΔN152-C28) and AvrBs3 (B)

To determine the active fraction of TALE variants a 5’-end Hex labeled specific DNA with a concentration (10 nM) much higher than the K_D values was applied. Labeling position of HEX is marked in yellow in the crystal structure of AvrBs3 bound to specific DNA (pdb: 2YPF) [205]. TALE variants were titrated with increasing concentration to the HEX labeled specific DNA. Each measuring point represents the mean ± standard error indicated by error bars for each point. n : number of independent experiments. Active fractions: 13% (AvrBs3(ΔN152-C28)) and 15% (AvrBs3(B)).

unlabeled DNA (Figure 38 (A)). HEX labeled specific DNA (10 nM) was incubated for 5 minutes with an excess of TALE (50 nM). Next, unlabeled-specific DNA or -scramble DNA was added to compete with HEX labeled specific-DNA (Figure 38 (B)). As a negative control, buffer was added instead of unlabeled specific/scramble DNA. Rate constant of AvrBs3 (B) for scramble DNA was \( k_1 = 9.6 \times 10^{-3} \pm 1.3 \times 10^{-3} \) s\(^{-1}\) with \( t = 103 \pm 11 \) s and rate constant for specific DNA was \( k_1 = 1.3 \times 10^{-3} \pm 4.2 \times 10^{-5} \) s\(^{-1}\) with \( t = 795 \pm 27 \) s (Table 21). Competition with unlabeled scramble DNA was fitted by a one-phase exponential decay function, whereas unlabeled specific DNA was fitted by two-phase exponential decay function.

Almost identical rate constants were determined by adding unlabeled scramble DNA and specific DNA consecutive to labeled DNA with an excess of AvrBs3 (B). Rate constant of AvrBs3 (B) for scramble DNA was \( k_1 = 9.7 \times 10^{-3} \pm 1.1 \times 10^{-3} \) s\(^{-1}\) with \( t = 102 \pm 11 \) s and rate constant for specific DNA was \( k_1 = 1.5 \times 10^{-3} \pm 1.1 \times 10^{-4} \) s\(^{-1}\) with \( t = 650 \pm 45 \) s.
Consecutively to the competition experiment, AvrBs3 (B) was digested by Proteinase K (PK) and baseline values of DNA with unbound protein were reached (Figure 39 (A); Table 21).

AvrBs3(ΔN152-C28) \((k_- = 2.1 \times 10^{-3} \pm 3.8 \times 10^{-3} \text{ s}^{-1}, t = 465 \pm 83 \text{ s})\) competed with unlabeled unspecific DNA 4.5 time slower, compared to AvrBs3 (B) \((k_- = 9.7 \times 10^{-3} \pm 1.1 \times 10^{-3} \text{ s}^{-1}, t = 102 \pm 27 \text{ s})\) (Figure 39 (B), Table 21). Competition to unlabeled specific DNA could not be fitted.

![Figure 38: Individual competition of AvrBs3 (B).](image)

(A) Crystal structure of AvrBs3 (pdb: 2YPF) [205] binding to 23 bp specific labeled DNA and unlabeled DNA. Labeling position of HEX is shown in yellow. (B) Individual competition: specific HEX labeled DNA was incubated with AvrBs3 (B) (50 nM). Then unlabeled specific DNA (180 nM; blue), unlabeled scramble DNA (180 nM; red), or buffer (green) was added to compete with specific DNA. Competition with unlabeled scramble DNA is fitted with a one-phase exponential decay function and competition with unlabeled specific DNA is fitted with a two-phase exponential decay function. \(n\) : number of independent experiments. Rate constant of AvrBs3 (B) with scramble DNA: \(k_- = 9.6 \times 10^{-3} \pm 1.3 \times 10^{-3} \text{ s}^{-1}\) with \(t = 103 \pm 11 \text{ s}\). Rate constant of AvrBs3 (B) with specific DNA: \(k_- = 1.3 \times 10^{-3} \pm 4.2 \times 10^{-5} \text{ s}^{-1}\) with \(t = 795 \pm 27 \text{ s}\).
Figure 39: Consecutive competition of AvrBs3 (B)

(A) Specific HEX labeled DNA was incubated with AvrBs3 (B) (50 nM) for 5 min (black), then unlabeled scramble DNA (180 nM) was added and fluorescence anisotropy changes were measured for 20 min (red). Thereafter, unlabeled specific DNA (180 nM) was added and fluorescence anisotropy was monitored over further 20 min (blue). At the end Proteinase K was added to digest the AvrBs3 (B) and the starting value was analyzed (magenta). Data points were fitted with a one-phase exponential decay function. Rate constant (B) Specific HEX labeled DNA was preincubated with AvrBs3 (B) (50 nM) for 5 min (black), then unlabeled scramble DNA (180 nM) was added and fluorescence anisotropy changes were measured for 20 min (red). Thereafter, unlabeled specific DNA (180 nM) was added and fluorescence anisotropy was detected over further 20 min (blue). Data points were fitted with an one-phase exponential decay function. Competition to unlabeled specific DNA could not be fitted.
Table 21: The $K_D$-values of AvrBs3($\Delta N152$-C28) and AvrBs3 (B)

The $K_D$-values of the TALE variants are listed for direct comparison. Considering the determined active fraction of 13% (AvrBs3($\Delta N152$-C28)) and 15% AvrBs3 (B) the $K_D$-values improve to $0.24 \pm 0.09$ nM (AvrBs3($\Delta N152$-C28)) and $0.14 \pm 0.05$ nM (AvrBs3 (B)) for 10% active TALE. Rate constants of dissociation for AvrBs3($\Delta N152$-C28) and AvrBs3 (B) for individual competition assay (1) and consecutive competition experiment (2) are also given.

<table>
<thead>
<tr>
<th></th>
<th>AvrBs3($\Delta N152$-C28)</th>
<th>AvrBs3 (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_D$ [nM]</td>
<td>1.9 $\pm$ 0.8</td>
<td>0.9 $\pm$ 0.3</td>
</tr>
<tr>
<td>$K_D$ [nM] 100% active protein</td>
<td>0.24 $\pm$ 0.09</td>
<td>0.14 $\pm$ 0.05</td>
</tr>
<tr>
<td>t [s]; $k^{-1}$ [s$^{-1}$] (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>scramble DNA</td>
<td>-</td>
<td>103 $\pm$ 11</td>
</tr>
<tr>
<td>specific DNA</td>
<td>-</td>
<td>795 $\pm$ 27</td>
</tr>
<tr>
<td></td>
<td>9.6 $\times$ $10^{-3}$ $\pm$ 1.3 $\times$ $10^{-3}$</td>
<td>1.3 $\times$ $10^{-3}$ $\pm$ 4.2 $\times$ $10^{-5}$</td>
</tr>
<tr>
<td>t [s]; $k^{-1}$ [s$^{-1}$] (2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>scramble DNA</td>
<td>465 $\pm$ 83</td>
<td>102 $\pm$ 11</td>
</tr>
<tr>
<td>specific DNA</td>
<td>2.1 $\times$ $10^{-3}$ $\pm$ 3.8 $\times$ $10^{-4}$</td>
<td>9.7 $\times$ $10^{-3}$ $\pm$ 1.1 $\times$ $10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>650 $\pm$ 45</td>
<td>1.5 $\times$ $10^{-3}$ $\pm$ 1.1 $\times$ $10^{-4}$</td>
</tr>
</tbody>
</table>

3.2.2.1 Electrophoretic mobility shift assay of AvrBs3 (AB)

Binding of AvrBs3 (AB) on a 5'-end Cy3 labeled 25 bp specific fragment was analyzed with an electrophoretic mobility shift assay (EMSA) on a 4% PAA gel. AvrBs3 was added with increasing concentration ranging from 0-600 nM to 5'-end Cy3 labeled specific DNA (20 nM) and mixtures were incubated for 5 min (0°C) (Figure 40). DNA total concentration was much higher then the $K_D$-value of the protein ([DNA]$_{total} >> K_D$). Thus, the $K_D$-value cannot be determined by this experimental set up. We expected to gain a curve, which increases linearly to its plateau. A plateau indicating the concentration, where all binding positions on the DNA are occupied. As expected, increasing protein concentration resulted in a linearly increasing curve (Figure 40 (B)).

Two DNA shifts have been detected by EMSA. The shift (I) increases linearly to a plateau of 0.46 $\pm$ 0.06 rel. fluorescence intensity, followed by a slight decrease due to the second shift. The shift (II) begins at a protein concentration between 250 nM and 450 nM and does not reach a plateau at 600 nM AvrBs3. In figure 40 (B), shift (I) and (II) are plotted and the rel. fluorescence intensity of shift (I) and (II) are added.
Figure 40: EMSA of AvrBs3 (AB)
AvrBs3 (AB) in increasing concentrations (0-600 nM) were preincubated for 5 min (0°C) with 20 nM 5’-end Cy3 labeled specific DNA. (A) EMSA was analyzed on a 4% PAA gel and the fluorophore Cy3 was excited and visualized with Typhoon™ FLA 9500 (532 nm). The 5’-end Cy3 labeled specific DNA is shown in green. Three distinct settings were detected, unbound DNA, shift (I) (blue), and shift (II) (red). (B) Relative fluorescence intensity [a.u.] of shift (I) (blue), shift (II) (red), and shift (I+II) (magenta) are plotted.
3.2.3 Labeled AvrBs3 variants

Since the mutated AvrBs3(ΔN152-C28) variants retained their binding characteristic, we analyzed if this is also the case for the labeled AvrBs3(ΔN152-C28) variants. The single cysteine variants were labeled with the fluorophore Cy5 as well as Alexa 488. Binding efficiencies were detected with anisotropy and FRET distance measurements. The double cysteine variant AvrBs3 (AB) was labeled (Alexa 488, Alexa 594, and Alexa 647) and binding characteristics were analyzed with EMSA.

3.2.3.1 Labeling of single cysteine variants with Alexa 488 fluorophore

First, the applicability of the fluorophore Alexa 488 for labeling AvrBs3 (B) was analyzed. The labeling reaction was performed with an excess of dye (1:4) for 1.30 h and resulted in a degree of labeling (DOL) of 158%. The binding characteristic was analyzed via anisotropy measurements on the 5'-end Hex labeled specific DNA (Figure 41). The $K_D$-value obtained for AvrBs3 (B)-Alexa 488 ($K_D = 1.3 \pm 0.3$ nM) was similar to the $K_D$-values of the unlabeled AvrBs3 variants.

![Figure 41: Determination of $K_D$-value of AvrBs3 (B)-Alexa 488](image)

Anisotropy measurement of AvrBs3 (B) labeled with the fluorophore Alexa 488 at the 5'-end Hex labeled DNA (0.5 nM). Protein was added with increasing concentrations to the 5'-end Hex labeled specific DNA. Each measuring point represents the mean ± standard error indicated by error bars for each point. n : number of independent experiments. Data points were fitted with a non-linear regression. $K_D$-value: 1.3 ± 0.3 nM.
3.2.3.2  Labeling of single cysteine variants with Cy5 fluorophore

Next, the single cysteine variants were labeled with the maleimide fluorophore Cy5. Labeling reactions were performed with an excess of dye (1:4) for 90 min (Figure 42 (A)). Labeled protein was separated from the free fluorophore using a Zeba Desalting column (40K) (Figure 42 (B)).

![Labeling of single cysteine variants with Cy5 fluorophore](image)

Figure 42: SDS-PAGE of single labeled AvrBs3 variants with fluorophore Cy5

Single cysteine variants of AvrBs3 were labeled with an excess of Cy5 maleimide fluorophore (1:4) and analyzed by 12% SDS-polyacrylamide gel electrophoresis. (A) Time resolved kinetics of the labeling reaction of AvrBs3 (A) (left) and AvrBs3 (B) (right). 2 μg protein was loaded in each lane. (-) represents the time point, when the fluorophore was added to the protein. The respective time points were taken (20’, 30’, 60’, 90’) and Cy5 fluorophore was excited at 635 nm with the Typhoon™ FLA 9500. (B) Images of the single cysteine AvrBs3 variants (4 μg) after purification with Zeba Desalting spin column (40K). Left: Coomassie Brilliant Blue stained 12% SDS-gel with Marker (PageRuler™ Unstained Protein Ladder; lane 1), AvrBs3 (A) (lane 2), and AvrBs3 (B) (lane 3). Right: Fluorescence image of the 12% SDS-gel visualized with Typhoon™ FLA 9500. The SDS-gel showed that AvrBs3 variants was analyzed above 85 kDa.
The Coomassie Brilliant Blue stained SDS-gel revealed that AvrBs3 variants run above 85 kDa (Figure 42 (B, left)). The SDS-gel can be visualized with fluorescence excitation (635 nm) (Figure 42 (B, right)). The degree of labeling of the single cysteine variants was analyzed. AvrBs3 (A)-Cy5 revealed a degree of labeling (DOL) of 115% and AvrBs3 (B)-Cy5 a DOL of 82%.

3.2.3.3  Förster resonance energy transfer (FRET) between Cy5 labeled AvrBs3 single cysteine variants and Cy3 labeled DNA

FRET enables a detailed analysis of the binding character of AvrBs3. The fluorophore Cy5 results in an effective FRET pair with fluorophore Cy3. Here, Cy3 served as a donor fluorophore and Cy5 as the acceptor fluorophore. The Förster radius of Cy3 to Cy5 is given with 60 Å [96]. Binding characteristics of the two Cy5 labeled single cysteine AvrBs3 variants were analyzed on two distinctively labeled 5'-end Cy3 specific DNAs (1 nM). The proteins were added to each DNA with increasing concentrations. Binding of the AvrBs3 variants to the DNA resulted in a donor quench and an increased indirectly acceptor signal (Figures 43, 54 and 55).

![Figure 43: Emission spectra of titrating Cy5 labeled AvrBs3 to 5'-end Cy3 labeled specific DNA (I)](image)

Representative example of emission spectra. AvrBs3 (A)-Cy5 analyzed on DNA (I) (1nM). Protein was titrated with increasing concentrations (0, 0.99, 2, 4, 6, 10, 14, 18 [nM]) to 1 nM DNA (I).
This FRET indexes (nF) were retrieved by measuring the acceptor emission resulted by the donor excitation (F_{\text{exD}emA}) and calculated by the three-cube method described by Youvan et al. [242]. This does not represent the FRET efficiency, but provides an index which change with FRET [17, 244]. The FRET signal had to be corrected via the crosstalk between the donor and the acceptor fluorophores. To receive the correction factors $\alpha$ and $\beta$ the fluorophores had to be excited both individual in the FRET channel ($F_{\text{exD}emA}$), the acceptor individual in the acceptor channel ($F_{\text{AexA}emA}$), and the donor individual in the donor channel ($F_{\text{DexD}emA}$). The donor was excited at 520 nm to reduce cross talk. The resulted spectra are given in figure 56 (appendix) for the donor fluorophore and figure 44 for the acceptor fluorophore. Furthermore, the cross talk considering the fluorescence signal $F_{\text{AexD}emA}$ and $F_{\text{DexA}emA}$ have to be analyzed. However, the FRET pair Cy3 and Cy5 resulted in a fluorescence signal $F_{\text{AexD}emA}$ and $F_{\text{DexA}emA}$ close to zero, thus this correction can be neglected in this experiment.

![Figure 44: Spectra of AvrBs3-Cy5 to determined the correction factor $\alpha$](image)

The correction factor $\alpha$ had to be defined to correct the crosstalk between Cy3 and Cy5. Acceptor fluorophore was excited at 520 nm (ex$^A$) and emission was detected in the acceptor channel as well as the acceptor was excited at 650 nm (ex$^A$) and emission was detected in the acceptor channel. The factor $\alpha$ is characterized by the fluorescence signal $F_{\text{AexD}emA}(A)$ divided with the fluorescence signal $F_{\text{AexA}emA}(B)$. The $F_{\text{AexD}emA}(A)$ indicates a partly direct excitation of the acceptor at 520 nm.

Both DNAs enclose 25 bp, but DNA (I) carries the Cy3 fluorophore at the 5’-end on the top strand (Figure 45 (A)) and DNA (II) is labeled at the 5’-end on the lower strand (Figure 46 (A)). The theoretical distances between Cy5 and Cy3 were measured on the crystal structure of AvrBs3 (pdb. 2YPF) [205]. The crystal structure contains a 23 bp long DNA, however in the experiment a 25 bp DNA was used. We added per missing bp 3.4 Å to the distance between Cy3 and Cy5. The measured distances $r_0^*$ are given in table 22, in figure 45 (A), and in figure 46 (A).
FRET measurements of AvrBs3 (A)-Cy5 with DNA (I) revealed a \(K_D\)-value of \(0.6 \pm 0.3\) [nM]. The FRET measurements of AvrBs3 (B)-Cy5 on DNA (I) also resulted in a comparable \(K_D\)-value of \(0.6 \pm 0.3\) [nM] (Figure 45 (B) and Table 22).

Figure 45: FRET distance measurement of AvrBs3 (A)-Cy5 and AvrBs3 (B)-Cy5 on DNA (I)
Single cysteine AvrBs3-Cy5 variants were analyzed via FRET distance measurements on 5′-end Cy3 labeled specific DNA (I) (1nM). (A) Crystal structure of AvrBs3 bound to 23 bp specific DNA [205]. 5′-end Cy3 labeled DNA (I) carried the fluorophore Cy3 (yellow) at the top strand 4 bp ahead of the target sequence. Both AvrBs3-Cy5 variants ((A) and (B)) were analyzed on DNA (I) resulting in a close distance (AvrBs3 (A)-Cy5: 41 Å) or a far distance (AvrBs3 (B)-Cy5: 87 Å) between Cy3 (yellow) and Cy5 (red). Theoretically calculated FRET efficiencies are given with E. (B) FRET index was calculated by formula (6) and then the corrected FRET signal (nF) was divided with the donor signal (Fex/Dem). Each measuring point represents the mean of FRET efficiency ± standard error indicated by error bars for each point. \(n\) : number of independent experiments. The data points were fitted with a non-linear regression curve. \(K_D\)-values: \(0.63 \pm 0.3\) [nM] (AvrBS3 (A)-Cy5); \(0.62 \pm 0.3\) [nM] (AvrBs3 (B)-Cy5).
Performing the FRET measurements with DNA (II) resulted in a $K_D$-value for AvrBs$_3$ (A)-Cy$_5$ of $1 \pm 0.6$ [nM], while for AvrBs$_3$ (B)-Cy$_5$ a $K_D$-value of $6 \pm 3$ [nM] was determined (Figure 46 (B) and Table 22).

![FRET distance measurement of AvrBs$_3$ (A)-Cy$_5$ and AvrBs$_3$ (B)-Cy$_5$ on DNA (II)](image)

Single cysteine AvrBs$_3$-Cy$_5$ variants were analyzed via FRET distance measurements on 5’-end Cy$_3$ labeled specific DNA (II) (1nM). (A) Crystal structure of AvrBs$_3$ bound to 23 bp specific DNA (pdb: 2YPF) [205]. 5’-end Cy$_3$ labeled DNA (II) carried the fluorophore Cy$_3$ (yellow) at the bottom strand 2 bp behind the target sequence. Both AvrBs$_3$-Cy$_5$ variants (A) and (B)) were analyzed on DNA (II), resulting in a far distance (AvrBs$_3$ (A)-Cy$_5$: $\sim58$ Å) or a close distance (AvrBs$_3$ (B)-Cy$_5$: $\sim40$ Å) between Cy$_3$ (yellow) and Cy$_5$ (red). Theoretically calculated FRET efficiencies are given with E. (B) FRET index was calculated by formula (6) and then the corrected FRET signal (nF) was divided with the donor signal (F$_{ex}$F$_{em}$D). Each measuring point represents the mean of FRET efficiency ± standard error indicated by error bars for each point. n: number of independent experiments. The data points were fitted with a non-linear regression curve. $K_D$-values: $1.1 \pm 0.6$ [nM] (AvrBS$_3$ (A)-Cy$_5$); $6 \pm 3$ [nM] (AvrBS$_3$ (B)-Cy$_5$).
The theoretical FRET efficiencies were calculated using the measured distance $r_0^*$ with the mathematic formula (9) and are listed in table 22. While it was expected to result a high FRET efficiency for the “close-state”, such as AvrBs3 (A) on DNA(I) ($r_0^* = 41 \text{ Å}$) and AvrBs3 (B) on DNA(II) ($r_0^* = 40 \text{ Å}$), a low FRET efficiency was expected for the “far-state”, such as AvrBs3 (B) on DNA(I) ($r_0^* = 87 \text{ Å}$) and AvrBs3 (A) on DNA(II) ($r_0^* = 58 \text{ Å}$).

Table 22: FRET distance measurements on DNA (I) and DNA (II)

Distance measurements between Cy3 and Cy5 were preformed in Py-MOL [195] and result in $r_0$ [Å]*. The experimentally determined $K_D$-values, the theoretically calculated FRET efficiencies (E) as well as the measured FRET efficiencies (E$_{fA}^*$) (mean values ± standard error for three measuring points with the highest protein concentration) are listed. The FRET efficiencies (E$_{fA}^*$) were calculated with formula (11). The FRET efficiencies (E$_{fA}^*$) were used to define $r_0$ [Å]** with formula (10).

DNA (I)

<table>
<thead>
<tr>
<th></th>
<th>$r_0^*$ [Å]</th>
<th>E</th>
<th>$K_D$-value [nM]</th>
<th>E$_{fA}^*$</th>
<th>$r_0^{**}$ [Å]</th>
<th>Eq</th>
</tr>
</thead>
<tbody>
<tr>
<td>AvrBs3</td>
<td>41</td>
<td>0.9</td>
<td>0.6 ±</td>
<td>0.42 ±</td>
<td>63</td>
<td>0.46 ±</td>
</tr>
<tr>
<td>(A)</td>
<td></td>
<td></td>
<td>0.3</td>
<td>0.03</td>
<td></td>
<td>0.02</td>
</tr>
<tr>
<td>AvrBs3</td>
<td>87</td>
<td>0.1</td>
<td>0.6 ±</td>
<td>0.52 ±</td>
<td>59</td>
<td>0.4 ±</td>
</tr>
<tr>
<td>(B)</td>
<td></td>
<td></td>
<td>0.3</td>
<td>0.06</td>
<td></td>
<td>0.1</td>
</tr>
</tbody>
</table>

DNA (II)

<table>
<thead>
<tr>
<th></th>
<th>$r_0^*$ [Å]</th>
<th>E</th>
<th>$K_D$-value [nM]</th>
<th>E$_{fA}^*$</th>
<th>$r_0^{**}$ [Å]</th>
<th>Eq</th>
</tr>
</thead>
<tbody>
<tr>
<td>AvrBs3</td>
<td>58</td>
<td>0.55</td>
<td>1 ± 0.6</td>
<td>0.69 ±</td>
<td>52</td>
<td>0.44 ±</td>
</tr>
<tr>
<td>(A)</td>
<td></td>
<td></td>
<td>0.12</td>
<td>0.06</td>
<td></td>
<td>0.06</td>
</tr>
<tr>
<td>AvrBs3</td>
<td>40</td>
<td>0.9</td>
<td>6 ± 3</td>
<td>0.74 ±</td>
<td>50</td>
<td>0.64 ±</td>
</tr>
<tr>
<td>(B)</td>
<td></td>
<td></td>
<td>0.06</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hoppe et al. [98] described a stoichiometric method based on the three cube-quantification that allows to measure FRET efficiency. We calculated the FRET efficiency (E$_{fA}^*$) for the plateaus with this formula (11). The resulted FRET efficiencies (E$_{fA}^*$) were used to recalculate the distance between Cy3 and Cy5 ($r_0^{**}$) via formula (10). Formula (10) is the result of converting formula (9) towards $r_o$. The distances $r_0^{**}$ are given in table 22. The calculations of FRET efficiencies (E$_{fA}^*$) on both DNAs revealed not the expected ratio of FRET efficiencies between “close-state” and “far-state” when compared to the theoretically calculated values (E).
The distance $r_{o}^{**}$ of AvrBs3 (A) on DNA(I) was $\sim$34% greater as the distance $r_{o}^{*}$ measured within the crystal structure, while AvrBs3 (A) on DNA(II) resulted in $\sim$10% reduced $r_{o}^{**}$ compared to the measured $r_{o}^{*}$. Similarly AvrBs3 (B) resulted in $\sim$32% reduced distance $r_{o}^{**}$ on DNA(I), while gaining $\sim$20% increase distance $r_{o}^{**}$ on DNA(II) when compared to the measured $r_{o}^{*}$ distance.

Furthermore, we calculated with formula (13) the donor quench for the plateau. It resulted in a similar donor quench for all constructs, except AvrBs3 (B)-Cy5 on DNA (II) which revealed an increased quenching effect.

3.2.3.4 Can AvrBs3 (AB) be labeled without affecting binding characteristics?

It has been discussed, that TALE binding proteins have a two-state search mechanism. On the one hand, the crystal structure revealed a looser conformation of unbound TALE protein, when compared to the bound state [57, 230]. On the other hand, a two-state mechanism was described recently in a single molecule study, distinguishing between a search state and a recognition state [51]. Additionally, the conformation elasticity of TALE was predicted with computational studies [226, 70]. Based on these observations, we aimed to reveal the conformational changes of TALE proteins with FRET distance measurements. To obtain a AvrBs3 variant which enable FRET distance measurements within the TALE protein, an active AvrBs3 variant carrying two distinct fluorophores had to be designed.

3.2.3.5 Fluorophores for labeling AvrBs3 (AB)

An AvrBs3 (AB) variant enclosing two cysteine residues, was labeled with single fluorophores to analyze the compatibility of the fluorophores and the binding character of TALE. AvrBs3 (AB) was labeled with the fluorophores Alexa 594, Alexa 488, or Alexa 647. Labeling reactions were performed for 2 h for each variant (Figure 47).

Labeling efficiency of AvrBs3 (AB)-Alexa 594 (Figure 47 (A), 48 ± 2%) and AvrBs3 (AB)-Alexa 488 (Figure 47 (B), 82 ± 2%) reached their plateau at approximately 60 min. AvrBs3 (AB) labeling reaction with Alexa 647 Figure 47 (C)) did not reach its plateau phase during the 2 h time period and achieved a labeling efficiency of 69%.
Figure 47: SDS-PAGE of AvrBs3 (AB) labeled with Alexa 594, Alexa 488, or Alexa 647
AvrBs3 (AB) was labeled with three distinct fluorophores: (A) Alexa 594, (B) Alexa 488, and (C) Alexa 647. The ratio of protein to dye corresponds to a 1:4 ratio. (-) represents starting point of the labeling reaction. Respective time points were taken during the labeling reaction (-, 15', 30', 60', 90', 120'). Labeling reactions were analyzed with 12% SDS-PAGE and fluorescent images were visualized with Typhoon™ FLA 9500. Each lane contained 2 μg protein.

Following the labeling reactions proteins were purified with a Zeba Desalting column (40 K) and DOLs were analyzed (AvrBs3 (AB)-Alexa 594: 91%; AvrBs3 (AB)-Alexa 488: 77%; AvrBs3 (AB)-Alexa 647: 144%). The binding characteristic of the single labeled AvrBs3 variants were detected with EMSA on a 5’-end Cy3 labeled specific DNA and are illustrated in figure 48 and figure 49. To ensure that the protein was located on the DNA with increasing concentrations, the fluorophore labeled to AvrBs3 (AB) was also exited on the corresponding 4% PAA gel. The two resulting pictures were merged and are displayed in figure 48. The 5’-end Cy3 labeled specific DNA is colored in green and the corresponding fluorophore is displayed in red. The analysis of DNA and protein results in a yellow spot. The merged pictures proofed that the labeled proteins did bind to the specific DNA (Figure 48 (B), (C) and Figure 49), except AvrBs3 (AB) labeled with Alexa 594 (Figure 48 (A) and Figure 49). The relative fluorescence intensities of the shifted DNA was plotted in figure 49 (B).
3.2 Characterization of Binding Characteristics of AvrBs3

Figure 48: Merge fluorescent images of AvrBs3 (AB) labeled with single fluorophores (Alexa 594, Alexa 488, or Alexa 647).

Merged fluorescence images of the 4% PAA gels were visualized with Typhoon™ FLA 9500. 5'-end Cy3 labeled DNA is excited at 532 nm and colored in green. Each fluorophore was exited with a specific wavelength and is shown in red. Region in which protein overlaps with DNA is displayed in yellow. (A) Merged fluorescence images of AvrBs3 (AB)-Alexa 594 using the green SHG laser (532 nm). Image was processed afterwards. No DNA shift is detected. (B) Merged fluorescence images of AvrBs3 (AB)-Alexa 488 were excited using the blue LD laser (473 nm). DNA is shifted, due to the binding characteristic of AvrBs (AB)-Alexa 488. (C) AvrBs3 (AB)-Alexa 647 was excited with the red LD laser (635 nm). Protein binding resulted in a DNA shift.
Figure 49: EMSA with AvrBs3 (AB) labeled with distinct fluorophores (Alexa 594, Alexa 488, and Alexa 647)
The protein binding was analyzed on a 25 bp long 5’-end Cy3 labeled specific DNA (20 nM, green) via EMSA. TALE protein was added in increasing concentrations to the 5’-end Cy3 labeled specific DNA (100, 150, 250, 450, 550, 600 nM). (A) The 4% PAA gels of the EMSA from AvrBs3 (AB)-Alexa 594 (I), AvrBs3 (AB)-Alexa 488 (II), and AvrBs3 (AB)-Alexa 647 (II). Fluorescence images were visualized with Typhoon™ FLA 9500 using the green SHG laser (532 nm). (B) Relative fluorescence intensity [a.u.] of the shifted DNA bands were plotted from each labeled AvrBs3 (AB) variant. AvrBs3 (AB)-Alexa 488 (red) and AvrBs3 (AB)-Alexa 647 (blue) bound to the specific DNA, but AvrBs3 (AB)-Alexa 594 (black) lost its binding characteristic.
3.2.3.6 Consecutive labeling of AvrBs3 (AB) with Alexa 488 and Alexa 647

Based on these results, AvrBs3 (AB) was labeled consecutively with Alexa 488 and Alexa 647 (Figure 50).

| + | + | + | + | Alexa 647 |
| - | + | + | + | Alexa 488 |
| 30' | 60' | 90' |

Figure 50: Double labeling of AvrBs3 (AB) with Alexa 647 and Alexa 488

Labeling reaction of AvrBs3 (AB) with Alexa 647 and Alexa 488 analyzed with a 12% SDS-PAGE. In each lane 2 μg protein was loaded. Merged fluorescence image was visualized with Typhoon™ FLA 9500, using the red LD laser (635 nm) and the blue LD laser (473 nm). Fluorescence signals: Alexa 647 (red) and Alexa 488 (green). Overlap of both fluorophores resulted in yellow color.

The labeling reaction was started with Alexa 647 for 30 min, then Alexa 488 was added to the reaction. The fluorophores were given to the reaction in a 1:2 ratio to the protein. After purification of the protein with Zeba Desalting columns (40K), the protein had a degree of labeling of 123% (Alexa 488) and 57% (Alexa 647). Next, the emission spectra of AvrBs3 (AB)-Alexa 488-Alexa 647 was detected, which indicated a FRET signal (Figure 51), thus a location proximity was proofed. The binding characteristic of AvrBs3 (AB)-Alexa 488-Alexa 647 was analyzed with EMSA, as described above (Figure 52). No DNA shift was detected.
Figure 51: Emission spectra of AvrBs3 (AB)-Alexa488-Alexa647

Alexa 488 was excited at 488 ± 5 nm using the Fluoromax-4 Spectrofluorometer. Emission spectra was detected between 500-720 nm. The donor fluorophore Alexa488 resulted in a maximum emission spectra at ~ 518 nm and the acceptor fluorophore Alexa 647 showed a peak at ~ 670 nm.

Figure 52: Binding characteristics of AvrBs3 (AB) - Alexa 484-Alexa 647

Binding was analyzed on a 4% PAA gel via EMSA. Protein was added with increasing concentrations (~ 100, 150, 250, 450, 550, 600 [nM]) to the 20 bp long 5′-end Cy3 labeled specific DNA (20 nM). The fluorescence images were visualized with Typhoon™ FLA 9500 using the green SHG laser (532 nm, Cy3), blue LD laser (473 nm, Alexa 488), and red LD laser (635 nm, Alexa 647). (A) 5′-end labeled Cy3 specific DNA. (B); (C): merged fluorescence images of excited DNA and each excited fluorophore. No DNA shift was detectable.
DISCUSSION

4.1 CHARACTERIZATION OF HIGHLY SPECIFIC NUCLEASES IN VITRO AND IN VIVO

Gene editing has been nominated to be the method of the year in 2011 by the journal Nature Methods [1]. It is a promising tool for curing monogenetic human diseases [50]. This can be realized using specific nucleases to disrupt a dysfunctional gene with NHEJ [131] or through gene replacement using an exogenous DNA template by the precise HDR [30, 95, 159, 60, 187]. In many monogenetic diseases gene replacement is required for efficient treatment [50]. However, HDR is less active than NHEJ and occurs only in the S/G2 phase of the cell cycle [43, 50]. In contrast, NHEJ occurs during the whole cell cycle [43, 50], but causes INDELs and chromosomal rearrangements [112, 16]. Furthermore, off-target effects are a major limitation for the successful application of gene editing. Off-target activity can induce unwanted chromosomal rearrangements, resulting in deletions, inversions, and translocations, thus causes cellular toxicity and long term consequences [104, 112]. Furthermore, there are concerns that off-target activity could activate oncogenes or inactivate tumor suppressors. Several types of engineered nucleases have been developed to optimize the efficiency and specificity of the cutting tools, such as monomeric TALE-FokI [210], TALE-PvuII [238, 239] or Cas9-FokI [89]. While even the most specific nucleases which induce DSBs have been reported to cause mutations at off-target sites [228, 93, 77], the newly designed Cas9-FokI seems to induce solely specific DSBs [89].

Specific nucleases inducing SSBs have been continuously optimized and represent promising tools for gene therapy [116, 157, 140, 78]. While the induction of SSBs increases HDR efficiency and reduces off-target effects, the overall efficiency of gene editing is reduced [145, 52, 222]. It has been recently reported that suppressing the key players of NHEJ increased efficiency of HDR by 4-5 fold [42]. Furthermore, it has been demonstrated that a nick in the transcribed strand triggers HDR eightfold more efficiently compared to a nick in the non-transcribed strand [53]. These findings encouraged us to analyze the site- and strand-specific nickase TALE-MutH under in vitro and in vivo conditions. Furthermore, studies with paired Cas9 nickases reported low off-target effects and genome editing efficiencies comparable to DSBs Schiml et al. [191], Shen et al. [199], Ran et al. [180], Cho et al. [40]. Thus, we aimed to adapt this double nicking approach to TALE-MutH and to optimize its specificity and efficacy.
TALE-MutH is a specific nuclease, which has been designed and analyzed under in vitro conditions by Gabsalilow et al. [78]. These studies revealed that TALE-MutH cuts either the lower strand or the upper strand depending on the target design. To optimize the therapeutic application of TALE-MutH we tested various target designs. Our data indicate towards the high therapeutic potential of TALE-MutH to enhance gene editing even when compared to the established specific nucleases I-SceI and TALE-FokI. To proof the therapeutical applicability of TALE-MutH, we exchanged the naturally occurring TALE binding protein AvrBs4 with engineered TALE binding domains, addressing the relevant gene target for X-linked retinitis pigmentosa which resulted in a robust activity.

4.1.0.7 AvrBs4-MutH

**Single targets:** The characteristics of TALE-MutH on the inverted single targets (T-3-H and H-3-T) were analyzed by employing an in vitro plasmid cleavage assay. TALE-MutH induced almost two times faster nicks to H-3-T, when compared to the single target T-3-H. However, analyzing the two single targets with the in vivo SSA assay led to no differences in the β-galactosidase (β-gal) activity. β-gal activity occurs due to the induced DNA breaks, which trigger SSA repair mechanisms, restoring the gene encoding the missing β-gal domain. Thus, the SSA assay did not confirm the finding by Davis et al. [53], that a nick in the transcribed strand triggers HDR more effectively, than a nick in the non-transcribed strand. However, the β-gal activity of the two single targets (T-3-H: 84 ± 9%; H-3-T: 84 ± 2%) were close to the I-SceI β-gal activity (I-SceI: 100 ± 1%). Following the analysis of the relative number of blue yeast colonies (Figure 22 (B)), the single targets T-3-H and H-3-T revealed a better activity than the β-gal activity of I-SceI. Additionally, we analyzed TALE-MutH on the single target T-6-H, which had been described by Gabsalilow et al. [78] to cut the top strand 3 times slower then the bottom strand on T-3-H under in vitro conditions. We were able to confirm a reduced activity of TALE-MutH on T-6-H by employing the in vivo SSA assay.

Interestingly, SSAs were triggered by inducing SSBs. While it is well recognized that HDR can be induced by SSBs with a low efficiency [145, 52, 222], the finding that SSAs can also be stimulated by SSBs has just recently been reported by Katz et al. [113].

These results indicated that TALE-MutH on single targets seem to be a promising tool for the in vivo application. Furthermore, when comparing the survival rate of TALE-MutH to the negative control (AvrBs4 without nuclease activity), no significant differences could be detected.

Analyzing the survival rate with the specific single targets only T-6-H revealed a significant reduction of the survival rate, when compared to the negative control. The survival rate of T-6-H proofed to be
4.1 CHARACTERIZATION OF HIGHLY SPECIFIC NUCLEASES in vitro and in vivo

comparable to that of the golden standard I-SceI and the inverted single targets revealed even an increased survival rate when compared to I-SceI. The latter finding is of high interest considering that TALE-MutH revealed a surprising high β-gal activity on the inverted single target.

**Double Targets:** TALE-MutH was analyzed on two double targets to investigate the double nicking approach of TALE-MutH. The first double target enclosed two adjacent GATC sites (T-3-HH-3-T). While adjacent GATC sites do not occur naturally, the experiment demonstrated the enzyme activity on a target without a distance between the two GATC sites. The second double target contained one GATC site flanked by the recognition sites of AvrBs4 (T-3-H-3-T). A pair of TALE-MutH introduced a site-specific DSB, thus creating a blunt end on T-3-HH-3-T and a 4 bp 5′overhang on T-3-H-3-T.

By using paired Cas9 nickases, it has been shown that the created 5′ overhang improves NHEJ and HDR efficiency to a comparable or even higher level when compared to induced blunt ends, while generating a 3′ overhang resulted in no improvement of the NHEJ and HDR efficiency [180, 140, 40]. Ran et al. [180] reported that no HDR was detectable with paired Cas9 nickases creating a 3′ overhang. Additionally, double nicking reduced the off-target effect by 50–1000 fold [180, 140, 40]. This encouraged us to analyze the characteristics of double nicking with TALE-MutH. The double targets were analyzed in vitro under denaturing conditions on a 5′ labeled PCR fragment. Here, TALE-MutH nicked T-3-H-3-T three to four times faster than T-3-HH-3-T. This finding indicated towards steric hindrance on the double target with two adjusted sites. Additionally, the in vitro kinetics revealed that TALE-MutH induces SSBs in a 1:1 ratio to top- and bottom-strand on both target sites. The question arise if TALE-MutH induced a DSB or only nicked the PCR fragments. Analyzing the 5′ labeled PCR fragments on native gels revealed that under in vitro conditions, no DSBs were induced unless TALE-MutH were digested with PK and new TALE-MutH were added. In this case, DSBs were detected particularly on the target T-3-H-3-T. When one TALE-MutH binds to its specific site, the target site is blocked under in vitro conditions.

The in vivo SSA assay revealed that β-gal activity on T-3-H-3-T (T-3-H-3-T: 99 ± 1%) is similar to the β-gal activity of I-SceI and the double target T-3-HH-3-T (T-3-HH-3-T: 95 ± 2%) is close to the β-gal activity of I-SceI. Both double targets showed a slightly higher β-gal activity, when compared to the inverted single targets. However, the relative number of blue colonies revealed a low activity on T-3-HH-3-T, whereas comparable levels to the inverted single target sites were found for T-3-H-3-T. Furthermore, the survival rates indicated that T-3-H-3-T is the preferred double target. Both targets showed a
significant reduction in the relative number of yeast colonies, when compared to AvrBs4 without nuclease activity, indicating a toxic effect on cell growth. Whereas T-3-HH-3-T had a lower survival rate compared to I-SceI, the survival rate of T-3-H-3-T was between I-SceI and the inverted single targets. Taken together, T-3-H-3-T proved to be an excellent target design combining high efficiency and low toxicity. However, it remains unclear if an optimized off-set between the two GATC sites in T-3-HH-3-T could improve SSA efficiency, as it has been reported for the paired Cas9 nickases [180]. The TALE-MutH design is limited by the availability of GATC sites in the genome and GATC sites close to each other are rare. Finally, the double target T-3-H-3-T revealed to be a suitable target design to induce a strand and site-specific DSBs with a efficiency comparable to I-SceI.

4.1.1 Beyond AvrBs4-MutH

4.1.1.1 AvrBs4-FokI

AvrBs4-FokI was the first TALEN construct and is based on the design of ZF-FokI [117]. The cleavage module consists of the non-specific cleavage domain of FokI. The cleavage domains have to dimerize to induce DSBs [167, 188, 23] and due to its non-specific nature high off-target effects have been reported [93]. We characterized TALE-FokI in our yeast assays, analyzing survival rates and β-gal activity. Comparing the survival rate without inducing the target plasmid, revealed no significant differences when compared to the negative control. However, analyzing the survival rates with target plasmids led to the highest reduction of the survival rate of all analyzed constructs, indicating towards toxic effects on cell growth.

The β-gal activity of AvrBs4-FokI (AvrBs4-FokI: 83 ± 17%) revealed a robust β-gal activity. Yet, the relative number of blue colonies showed the toxic effects of AvrBs4-FokI. Both, I-SceI and TALE-MutH on all tested targets displayed a higher specificity and efficiency, when compared to TALE-FokI. In table 23 the in vivo characteristics of I-SceI, TALE-FokI and TALE-MutH are summarized.

4.1.1.2 AvrBs3-PvuII

An alternative architecture of the specific TALENs is the combination of the site-specific cleavage domain PvuII with the binding module TALE [238, 239]. Our data showed that the TALE-MutH variant has superior characteristics for gene editing, when compared to the TALE-PvuII variants. PvuII was initially used as an alternative cleavage domain to FokI in combination with the binding module TOF [65]. Similar to FokI, PvuII has to dimerize to induce DSBs. The main advantage of PvuII is its site-specific character, which is based on the recognition of a palindromic 6 bp sequence [37, 170, 169]. Due
to the limitation of TOF-sc-PvuII for in vivo applications the specific nuclease were optimized using binding modules such as ZFs or TALEs [239, 190]. Previous, investigations reported that TALE-PvuII had superior features, when compared to ZN-PvuII [190]. Today, three TALE-PvuII variants are available (TALE-hf-PvuII, TALE-sc-PvuII and TALE-hd-PvuII). The variants were designed and optimized by Yanik et al. [239, 238]. Using the two distinct yeast viability assays, we were able to confirm the high toxicity of TALE-hf-PvuII and TALE-sc-PvuII, which has been previously reported by Yanik et al. [239, 238] in E.Coli strains. To optimize the TALE-PvuII design, a heterodimeric variant was designed by introducing a mutation in each subdomain of PvuII [238]. Under in vitro conditions, the variants AvrB33-K38D-PvuII and AvrB33-D30K-PvuII have been shown to be inactive on their own, but when introduced together they induce specific DSBs. Similar results were achieved when the variants were analyzed in HEK293 cells, however with a reduced activity [238]. We could not reproduce this finding in yeast cells. Here, isolated AvrB33-D30K-PvuII showed no activity, while AvrB33-K38D-PvuII (β-gal activity: 95 ± 3%) revealed a similar β-gal activity as I-SceI. The relative number of blue colonies of AvrB33-K38D-PvuII was significantly higher, when compared to the negative control. However, it revealed a reduced activity, when compared to I-SceI. Analyzing the
heterodimeric AvrBs3-hd-PvuII combination revealed a reduced activity, when compared to AvrBs3-K38D-PvuII. Thus, it can be suggested that AvrBs3-K38D-PvuII is active in yeast.

The viability assay without a specific target revealed no significant difference in the relative amount of cell survival between the negative control and the analyzed TALE-PvuII variants. However, the viability assay with specific targets showed a reduced cellular growth for AvrBs3-hd-PvuII and AvrBs3-K38D-PvuII.

Taken together, the different AvrBs3-PvuII variants proofed to possess major limitations in the SSA assay and the viability assays. We were able to confirm the high toxicity of TALE-hf-PvuII and TALE-sc-PvuII with our SSA assay, but the reported function of the heterodimeric AvrBs3-hd-PvuII could not be reproduced. Contrasting to the reported characterization of AvrBs3-K38D-PvuII [238], isolated AvrBs3-K38D-PvuII revealed to be able to induce DSBs in yeast. An overview of the characteristics of AvrBs3-variants is given in table 24.

### Table 24: Overview of the nuclease activity of AvrBs3-PvuII variants

Significant difference between analyzed AvrBs3-PvuII variants and AvrBs4 without nuclease activity were determined by the unpaired T-test: n ≥ 3, n.s. = not significant, P < 0.05 *, P < 0.01 **, P < 0.001 ***

<table>
<thead>
<tr>
<th></th>
<th>Survival rate</th>
<th>β-gal activity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>with target site</td>
<td>without target site</td>
</tr>
<tr>
<td>I-SceI</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>AvrBs3-hf-PvuII</td>
<td>*</td>
<td>n.s.</td>
</tr>
<tr>
<td>AvrBs3-K38D-PvuII</td>
<td>*</td>
<td>n.s.</td>
</tr>
<tr>
<td>AvrBs3-D30K-PvuII</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>AvrBs3-sc-PvuII</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td>AvrBs3-hd-PvuII</td>
<td>*</td>
<td>***</td>
</tr>
</tbody>
</table>

4.1.1.3 *Cas9 endonuclease*

The expression plasmid for Cas9 endonucleases in yeast has been established in 2013 [58] and we aimed to investigate the β-gal activity of Cas9 endonucleases with the SSA system. We were able to confirm
a significant induction of β-gal activity in *S. cerevisiae* by the Cas9 endonuclease, when compared to the negative control. Still, compared to I-SceI the Cas9 system showed low β-gal activity. Three plasmids had to be introduced; the expression plasmid of Cas9, the gRNA plasmid, and the target plasmid. The expression plasmid for Cas9 nuclease and the target plasmid carry the same selective marker (TRP1). Thus, our results could be significantly influenced by transformation efficiency.

4.1.1.4 *Alignment search of the addressed sites in the genome of Saccharomyces cerevisiae*

We screened the yeast genome of the S288c strain for similar off-target sites of the employed specific nucleases. No 100% similar off-target site for the analyzed on-target sites were found. Next, the seed regions of the target site for each specific nuclease was analyzed. AvrBs4 and AvrBs3 revealed the highest off-target similarity with 68% containing the 5’ seed region. This finding is in agreement with the low cell growth of AvrBs4-FokI. Yet, no similar off-target was found including the 10 bp of the 5’ seed region and the addressed site of MutH (5’-GATC-3’) or PvuII (5’-CAG CTG-3’). For the on-target of I-SceI, the best two hits had a similarity of 66%, enclosing the 3’ seed region of I-SceI. The on-target for Cas9::gRNA revealed off-target site hits with a similarity of 52% containing the PAM site and the 3’ seed region. Cas9::gRNA has been shown to possess a lower specificity, compared to the TALE binding domains. Furthermore, the Cas9 system has been reported to tolerate up to 5 mismatches [101, 163, 76]. It is therefore, that the difference of the off-target sites and on-target sites are adequate to limit off-target effect.

**Therapeutic target:** The next step was to analyze if the naturally occurring binding module AvrBs4 could be exchanged by an engineered TALE binding domain without losing the specificity and efficiency of the designed nuclease TALE-MutH. We targeted the relevant gene in X-linked retinitis pigmentosa, which is caused by mutations in the RPGR-gene. 80% of the mutations leading to RPGR dysfunction are located in the ORF15, which consists primarily of an imperfect repeat motif [100, 223, 198, 165]. Treatment of X-linked retinitis pigmentosa could be realized by insertion of a AAV vector enclosing the encoding sequence for the RPGR as has been demonstrated in dogs [15] and mice [234].

Distinct GATC sites are located in the ORF15 and TALE binding modules were created by employing the Golden Gate cloning strategy. The newly designed TALE-MutH constructs (TALE(RPGR)-1 (I)-MutH/ TALE(RPGR)-2 (I)-MutH) were analyzed in terms of their activity and toxicity in HEK293T cells. They addressed a heterodimeric double target site containing one GATC site flanked with the two
TALE binding sites. The design is based on the double target T-3-H-3-T and the heterodimeric TALE-MutH induces a strand- and site-specific DSB resulting in a 4 bp 5’overhang. The activity was detected with a frame shift assay, which is triggered by NHEJ. The high frequencies of NHEJ and the independence of a DNA template simplify the analysis of nuclease activity, when compared with the detection of HDR efficiency. Thus, we analyzed NHEJ activity to characterize nuclease activity on the therapeutic target.

It revealed similar NHEJ activity for the heterodimeric TALE-MutH design (TALE(RPGR)-1 (I)-MutH/ TALE(RPGR)-2 (I)-MutH), when compared with the Golden Standard I-SceI. Additionally, no significant difference between the survival rates of I-SceI and the heterodimeric TALE-MutH constructs were found. Furthermore, an empty plasmid (pcDNA) was analyzed as a negative control. No significant difference between the survival rates of I-SceI and empty vector were detected, while the ZF-FokI [6] construct revealed a reduced survival rate. The episomal in vivo analysis of TALE-MutH on the therapeutic target underlined its high potential as a specific and efficient tool for gene targeting.

Our results indicate that TALE-MutH can challenge the established specific nucleases (I-SceI and TALE-FokI), when addressing the relevant gene causing 80% of X-linked retinitis pigmentosa [100, 223, 198, 165]. The target site T-3-H-3-T, proofed to be a powerful target design to induce double nicks. Further important steps are the analysis of HDR efficiency of TALE-MutH on the T-3-H-3-T targets, charactering the exogenous DNA template, such as its overall length, the length of the homologous parts, or its double stranded or single stranded DNA character. Additionally, the effect of TALE-MutH addressing its target site on a genome have to be investigated. Here, the chosen genome should be screened for similar off-target sites and the accessibility of the target sequence in the chromatin structure remains to be analyzed.

Nevertheless, TALE-MutH requires a GATC site to induce a SSBs and TALE binding modules have to be designed for each new target site respecting several conditions, such as the required thymine at the beginning of the target sequence [193] and the higher sensitivity to mismatches at the 5’end of the addressed site [142].

The Golden Gate cloning strategy enables to design TALE binding modules for any given target within a 5 days protocol [67]. TALE binding modules can be very specific [215, 160, 59], but depending on the addressed target site, the TALE domain can also looses its specific features [88]. Computer programs have been developed to predict the best composition of a TALE binding domain [184, 86, 62]. Yet, an improved understanding of the TALE binding mechanism remains to be investigated.
The second part of this thesis concentrates on the characterization of the naturally occurring TALE protein AvrBs3. Four distinct AvrBs3(ΔN152-C28) variants (two single cysteine variants, a double cysteine variant, and a variant containing all cysteine residues) were expressed and characterized. The $K_D$-values of the AvrBs3 variants were analyzed and functional labeled AvrBs3 variants were designed.

TALE proteins have been used as binding module for specific nucleases and were successfully employed in genome and epigenome editing technologies [109, 48, 135]. Deciphering of the recognition code of TALE [25] and the introduction of the one pot cloning strategy GOLDEN GATE represent important steps toward a therapeutic use of the TALE proteins [33]. Theoretically, any given target can be addressed by an engineered TALE protein. Despite these developments, off-target effects of TALENs were described and limit their application [88]. Improving the knowledge on the binding mechanism of TALE has been a central aim of many researchers [142, 205, 193, 194]. For instance, a polarity effect has been recognized, indicating that the 5’-end is more sensitive to mismatches [142]. Furthermore, the impact of adjacent repeat domains were investigated and computer programs were developed to identify the best fitting TALE construct for a specific target [184]. Still, a better understanding of the binding mechanism of TALE proteins is required, which could result in an improved prediction of potential off-target effects.

4.2.0.5 Determination of $K_D$-values

The $K_D$-values of AvrBs3 with different lengths of its N- and C-terminus were analyzed under various conditions using several methods [205, 142, 194]. Schreiber et al. [194] described the $K_D$-value of AvrBs3(ΔN152-C16) with similar N-terminus length to AvrBs3(ΔN152-C28) using fluorescence anisotropy on a 36 bp long DNA under physiological conditions (c(NaCl) = 150 mM, pH: 7.5). Their studies revealed a higher $K_D$-value of the AvrBs3(ΔN152-C16) ($K_D = 282 \pm 18$ nM), when compared to the $K_D$-value of the AvrBs3(ΔN152-C28) ($K_D = 1.9 \pm 0.8$ nM) as we have determined on a 25 bp long DNA under low salt conditions (60 mM). These divergent results could be explained by the extended DNA and the high salt concentration employed by Schreiber et al. [194]. Additionally, they measured the $K_D$-values of wtAvrBs3 ($K_D = 23.2 \pm 1.6$ nM) and AvrBs3(ΔN152-C278) carrying the entire C-terminus ($K_D = 99.2 \pm 7.5$ nM), suggesting that a truncation of the C-terminus to 16 aa has a high impact on the binding characteristics of AvrBs3. Stella et al. [205] measured the $K_D$-value of wtAvrBs3 ($K_D = 33.6 \pm 4.9$ nM) with fluorescence anisotropy on a 20 bp long DNA under physiological conditions (c(NaCl) = 150 mM, pH: 7.5). The $K_D$-values of wtAvrBs3 in both studies were compa-
rable. However, using the same conditions, but isothermal titration calorimetry for detection resulted in a $K_D$-value of $88 \pm 32$ nM [205].

A third group reported $K_D$-values of $3.2 \pm 1.6$ nM (AvrBs3(Δ N143-C54)) and $3.9 \pm 0.1$ nM (AvrBs3(Δ No-C192)) using EMSA under low salt conditions (c (KCl)= 60 mM) [142]. These findings were similar to the $K_D$-value of AvrBs3(Δ N152-C28) ($K_D= 1.9 \pm 0.8$ nM) measured in our study. Furthermore, the ratio of active and inactive AvrBs3 influences the $K_D$-value. Here, protein expression, purification, and storage have an impact on the ratio between active and inactive protein and could be important factors for comparing the analyzed $K_D$-values. We determined the ratio of active AvrBs3(Δ N152-C28) to estimate the $K_D$-value for 100% active protein. About 13% of AvrBs3(Δ N152-C28) was active, resulting in a $K_D$-value of $0.24 \pm 0.09$ nM. The high inactive portion of AvrBs3(Δ N152-C28) indicates towards a relevant degeneration under in vitro conditions. An activity assay with the fusion construct AvrBs3-PvuII revealed a 50% reduction of the cutting activity under physiological conditions in 4 h [238].

**4.2.0.6 Cysteine variants**

The characterization of the crystal structure of the apoenzyme and the bound TALE protein was a major step in identifying the binding character of the TALE::DNA interaction. Here, a superhelical conformation of TALE was shown to interact with the major groove of the DNA [139, 205] and a positive charge along the inner surface contributes to the non-specific binding energy of the TALE::DNA interaction [57, 81]. The positive charged amino acids G14, K16 and Q17 interact with the negatively charged phosphate backbone of the DNA, thus supplying the main binding energy. Interestingly, the so called di-residues at position 12 and 13 do not significantly contribute to the binding energy, despite the fact that residue 13 is the only known AA performing direct interactions with the DNA. This suggests, that binding to the specific DNA occurs by negative discrimination due to steric and electrostatic clashes between DNA and the di-residues [230, 57, 139]. Furthermore, the crystal structure and computer based analysis of TALE binding revealed a conformation elasticity of the binding module, resulting in a loose conformation as apoenzyme and a more tight conformation when bound to its target site [57, 230, 226, 70]. It has been reported, that the superhelical turn of TALE is reduced from 60 A° to 35 A° in its bound state [57]. Recently, the AvrBs3 binding mechanism has been described on the single-molecule level [51]. To enable this, AvrBs3 was labeled with Cy3 at the N-terminus via an unnatural aldehyde. The TALE protein was tagged with a six AA motif and the formylglycine generating enzyme (FGE) converted the cysteine enclosed in the motif on protein level. The single-molecule experiment revealed a conformation dynamic between search and recognitions state, as well as the interactions between NTR and non-specific DNA
that facilitate 1D diffusion [51]. The main advantage of the labeling strategy used by Cuculis et al. [51], is the lack of mutations which could impact the binding character of AvrBs3. However, AvrBs3 hold a tag at the N-terminus, which could influences the binding characteristics due to steric clashes. Moreover, the fluorophore is fused to the flexible linker and therefore precise distance measurements via FRET are not possible. Another strategy to label proteins is to create single cysteine variants. While the fluorophore is located closer to the protein, the cysteine residue has to be accessible for the labeling reaction and the removed cysteine residues are not allowed to have a mayor impact to the binding characteristics of the protein. AvrBs3 has a cysteine residue at position 30 in each repeat domain [57, 139]. In AvrBs3(ΔN152-C278) Schreiber et al. [194] substituted all cysteine residue to serine or alanine residues (AvrBs3(C30S)Rep / AvrBs3(C30A)Rep). This study proofed that the cysteine residues are required for gene activation in planta. Interestingly, the specific interaction between AvrBs3(C30S)Rep::DNA is only minimally reduced under in vitro conditions, when compared to AvrBs3(ΔN152-C278) containing all cysteine residues. The cysteine residues are required for dimerization via disulfide bridges to homodimers, which facilitate the nuclear transport [91, 194]. AvrBs3(C30S)Rep revealed reduced complex formation to AvrBs3(ΔN152-C278) due to the missing cysteine residues. Furthermore, Schreiber et al. [194] postulated that cysteine substitution resulted in a plasticity change of AvrBs3. The missing cysteine residues led to a small conformation variance in each repeat. This effect might increase with increasing repeat numbers and result in the incapability of AvrBs3(C30S)Rep to activate genes in planta.

In the present thesis, the cysteine residue was substituted to methionine, as methionine has the least shift to cysteine in its thermodynamic stability (ΔΔG = 0.3 [kcal/mol]) when compared to all other aa. Serine and alanine revealed a shift in the thermodynamic stability above 1 [kcal/mol]. We designed three distinct cysteine variants. AvrBs3 (A) carrying one cysteine residue in repeat one, AvrBs3 (B) encloses a cysteine residue in repeat 16, and AvrBs3 (AB) contains two cysteine residues. First, we analyzed the K_D-value of AvrBs3 (B) and were able to show that the K_D-value of AvrBs3 (B) was similar to the K_D-value of AvrBs3(ΔN152-C28) containing all cysteine residues. Consequently, the cysteine residues did not affect the binding characteristics of AvrBs3 under in vitro conditions regarding the determined K_D-values. While the substitution of cysteine to methionine revealed similar binding characteristics compared to AvrBs3(ΔN152-C28), the substitution described by Schreiber et al. [194] resulted in a slightly decreased affinity of AvrBs3(C30S)Rep compared to AvrBs3(ΔN152).

Next, we performed competition assays in which AvrBs3 (B) bound to its specific target site (AvrBs3 (B)::DNA) competed with an ex-
cess of non-specific or specific free DNA. The experiment revealed a non-specific binding state and a specific binding state. This finding was supported by an individual competition assay and a consecutive competition assay with AvrBs3 (B). Our results suggest, that several AvrBs3 (B) are clustered to the specific AvrBs3 (B)::DNA complex. Adding an excess of unspecific DNA to the specific AvrBs3 (B)::DNA complex resulted in a fast displacement of the non-specific bound proteins, followed by a slow displacement of the specific interaction (AvrBs3 (B)::DNA) when an excess of specific DNA was added. Removing all specific interactions was achieved by digesting AvrBs3 (B) with Proteinase K.

On the contrary, AvrBs3(ΔN152-C28) carrying all cysteine residues did result in an even slower decrease of the specific interaction (AvrBs3 (ΔN152-C28)::DNA) by adding an excess of specific DNA to the specific complex. These findings indicate towards a stronger AvrBs3(ΔN152-C28)::DNA interaction, when compared to the cysteine substitution variant (AvrBs3 (B)::DNA).

4.2.0.7 Labeling single cysteine AvrBs3 variants

Next, we labeled the AvrBs3 cysteine variants and analyzed their binding characteristics under in vitro conditions. AvrBs3 (B) labeled with Alexa 488 resulted in a DOL of 158% and a $K_D$ value of $1.3 \pm 0.3$ nM. This proofed that the fluorophore Alexa 488 is compatible with the binding characteristics of AvrBs3. The $K_D$-value is comparable with AvrBs3(ΔN152-C28).

AvrBs3 (A) and AvrBs3 (B) were also labeled with the Cy5 fluorophore and DOLs of 115% (AvrBs3 (A)-Cy5) and 82% (AvrBs3 (B)-Cy5) were achieved. The binding characteristics of the variants were analyzed on two distinct 25 bp long 5'-end Cy3 labeled specific DNAs. DNA(I) carried the Cy3 fluorophore at the 5'-end on the top strand and DNA(II) held the Cy3 fluorophore at the 5'-end on the lower strand. Cy5 and Cy3 are a known FRET pair with a FÖRSTER RADIUS of 60 Å [96]. The FRET measurement depends on the overlapping absorption spectra of donor and acceptor fluorophores, the distance between the chosen fluorophores and the orientation between the donor and acceptor fluorophores to each other.

The $K_D$ -values of the different variants were analyzed on the two DNA designs with FRET measurements. The measured FRET signals were corrected by the cross-talk between donor and acceptor fluorophores. The FRET indexes (nF) were calculated by the three-cube method described by Youvan et al. [242]. This method does not represent the FRET efficiency, but provides an value, which changes with FRET [17, 244], thus it enables determination of $K_D$ -values.

On DNA(I) both Cy5 labeled AvrBs3 variants revealed comparable $K_D$ -values to the unlabeled single cysteine variant. On DNA (II)
AvrBs3 (A)-Cy5 also showed a similar K_D -value when compared to the unlabeled single cysteine variant, while AvrBs3 (B) revealed an increased K_D -value of 6 ± 3 nM, when compared to the anisotropy results of the unlabeled cysteine variants.

The binding character of AvrBs3 was studied by using the specifically labeled AvrBs3 variants on DNA(I) and DNA(II). AvrBs3 (A)-Cy5 on DNA(I) (r_0* = 41 Å) and AvrBs3 (B) on DNA II) (r_0* = 40 Å) was expected to have a “close-state” to the Cy3 label on the DNAs leading to an increased FRET efficiency, when compared to AvrBs3 (B)-Cy5 on DNA(I) (r_0* = 58 Å) and AvrBs3 (A)-Cy5 on DNA(II) (r_0* = 87 Å), representing the “far-state”.

We measured the theoretical distance between donor and acceptor fluorophores within the crystal structure of AvrBs3 on a 23 bp long DNA (pdb: 2YPF [205]). The DNAs used in this thesis were 25 bp long and therefore 3.4 Å were added for each missing nucleotide. The theoretical FRET efficiencies were calculated with the measured distance between donor and acceptor fluorophores using formula (9). These data revealed that the difference between “close-state” and “far-state” is more pronounced on DNA(I) (E(AvrBs3 (A))= 0.9; E(AvrBs3 (B))= 0.1), compared to DNA(II) (E(AvrBs3 (A))= 0.55; E(AvrBs3 (B))= 0.9). Additionally, the difference between the theoretically calculated FRET efficiencies of AvrBs3 (B)-Cy5 on the two DNAs is larger as the observed FRET efficiencies of AvrBs3 (A)-Cy5 on the two different DNAs.

Furthermore, we calculated the FRET efficiency (E_f^*) for the three measuring points with the highest protein concentration with a stoichiometric method that is based on the three cube-quantification described by Hoppe et al. [98]. The resulted FRET efficiencies (E_f^*) were used to recalculate the distance between Cy3 and Cy5 (r_0**) as described by formula (11).

The measured FRET efficiencies (E_f^*) on both DNAs revealed not the expected ratio of FRET efficiencies between “close-state” and “far-state” when compared to the theoretically calculated values (E). The distance r_0** of AvrBs3 (A) on DNA(I) was ~34% greater as measured within the crystal structure, while AvrBs3 (A) on DNA(II) resulted in ~10% reduced r_0** compared to the measured r_0*. Similarly AvrBs3 (B) resulted in ~32% reduced distance r_0** on DNA(I), while gaining ~20% increase distance r_0** on DNA(II) when compared to the measured r_0* distance (Figure 53). These finding indicates towards diverged orientation of the Cy3 and Cy5 fluorophore as expected. The fluorophores were covalent bound to the cysteine residue allowing a certain elasticity of the fluorophores. Due to the plastic elasticity of AvrBs3 and the fluorophore Cy5 the location of Cy5 fluorophore was hardly described. A further step would be the calculation of an expected flexibility of Cy5 according to its structural size.
The resulted distances $r_{0}^{**}$ indicated that the fluorophore Cy5 labeled to the N-terminus of AvrBs3 (A) is shifted to the C-terminus, as well as the Cy5 fluorophore labeled to the C-terminus (AvrBs3 (B)) is shifted to the N-terminus. These would explain the greater distances for the “close-state” and the reduced distances for the “far-states” (Figure 53).

The $E_{fA}^{*}$ were calculated by the formula (11) described by Hoppe et al. [98]. However, Hoppe et al. [98] also described an extended formula which corrects the FRET signal for the donor quench, thus the FRET efficiencies could be defined more accurate. Additionally, FRET signal could also be influenced by different factors which complicate the comparison between the single experiments. For instance, the different DOLs of the labeled DNA and AvrBs3-Cy5, further quenching effects, and preferred orientation of Cy3 and Cy5 within one of the DNA::AvrBs3 complexes when compared with another DNA::AvrBs3 complex.
Quenching in term of FRET has also become a tool to investigate ligand binding, thus we calculated the donor quench for the plateau. It resulted in a similar donor quench for all constructs, except AvrBs3 (B)-Cy5 on DNA (II) which revealed an increased quenching effect. This finding indicates towards similar binding characteristics of the AvrBs3::DNA complexes.

4.2 Characterization of Binding Characteristics of AvrBs3

4.2.0.8 Labeling of the double cysteine variant AvrBs3 (AB):

It has been described, that TALE has a conformational elasticity. Binding to DNA results in a tighter superhelical conformation [57, 230, 226, 70], but conformational changes between non-specific and specific binding has also been described with single-molecule experiments [51]. We aimed to show this effect with FRET distance measurements. We designed a double cysteine AvrBs3 (AB) variant, carrying a cysteine residue in repeat 1 and in repeat 16. We expected, that upon binding to the DNA the distance between donor and acceptor fluorophore would decrease, resulting in an increased FRET signal.

The first challenge was to design a functional double labeled AvrBs3 variant. The binding characteristics of unlabeled AvrBs3 (AB) was analyzed by EMSA, which revealed two shifts of the labeled DNA. This suggested that the first shift contributes to the specific binding and the second shift is a result of non-specific binding clustered to the specific AvrBs3 (AB)::DNA complex. The next step was to find donor and acceptor fluorophores compatible with the binding characteristics of AvrBs3 (AB). AvrBs3 (AB) was labeled either with Alexa 488, Alexa 594, or Alexa 647; EMSA was performed to analyze the binding capacity. AvrBs3 (AB) labeled with Alexa 488 and Alexa 647 kept its binding characteristics, but AvrBs3 (AB)-Alexa 647 showed decreased binding to the specific DNA. AvrBs3 (AB)-Alexa 594 completely lost its binding characteristics. Alexa 488 and Alexa 647 have overlapping absorption spectra and function as a FRET pair. Therefore, AvrBs3 (AB) was labeled consecutively with Alexa 488 and Alexa 647. Based on the isolated labeling kinetics with Alexa 488 and Alexa 647, we knew that Alexa 647 reacts slowly with the cysteine residue of AvrBs3 (AB). Therefore, AvrBs3 (AB) was incubated for 30 min with Alexa 647 and then Alexa 488 was added. Despite this, a DOL for Alexa 488 of 123% and for Alexa 647 of 57% were measured. Thus, Alexa 647 showed a reduced labeling capacity for AvrBs3 (AB). Alexa 488 (M = 721 [g/mol]) has a small molecular weight, when compared to Alexa 647 (≈ 1250 [g/mol]). This could result in steric hindrance, which causes the reduced labeling capacity of Alexa 647 to AvrBs3 (AB).

Next, we analyzed if AvrBs3 (AB)-Alexa 488- Alexa 647 kept its binding characteristics by employing EMSA. However, it revealed a complete loss of the protein::DNA interaction. Thus the FRET pair
Alexa 488 and Alexa 647 could not be used for this experimental approach. We could proof that AvrBs3 (AB)-Alexa 488- Alexa 647 displayed a FRET signal. Consequently, the donor and acceptor fluorophores were close to each other, which indicates towards an aggregation of the protein due to the covalently bound labels.

Further investigations identifying compatible FRET pairs with AvrBs3 (AB) binding capacity are required. For instance, the FRET pair Cy3 and Cy5 could be analyzed, but also Cy5 in combination with Alexa 488 could be a promising attempt, as Alexa 488 has shown an advantageous labeling capacity.

In conclusion, we analyzed the in vitro binding characteristics of AvrBs3(Δ N152-C28) under low salt conditions and revealed a similar KD-value to the reported value of Meckler et al. [142]. The other groups determined the binding characteristics under physiological conditions and measured higher KD-values [194, 205]. We were able to design two single cysteine variants and one double cysteine variant, which kept their in vitro binding characteristics. The competition assays revealed a stronger protein::DNA interaction of AvrBs3(Δ N152-C28), when compared to the single cysteine variant. It remains unclear, if the substitution of cysteine to methionine improves gene activation under in vivo condition, when compared to the AvrBs3(C30S)Rep designed by Schreiber et al. [194].

Labeled single cysteine variants with Alexa 488 and Cy5 kept their binding characteristics and could be used to further investigate the binding mechanisms of the AvrBs3 in vitro.

The FRET efficiencies measurements (EfA*) with AvrBs3 (A)-Cy5 and AvrBs3 (B)-Cy5 on DNA(I) and DNA(II) were not in concordance with the expected “close-state” and “far-state”. The results indicated towards a displacement of the Cy5 fluorophore when compared to the expected location within the crystal structure.

Finally, a double cysteine AvrBs3 labeled with a donor and an acceptor fluorophore was designed. Isolated labeled AvrBs3 (AB) with Alexa 488 or Alexa 647 kept its binding capacity, while consecutively labeling resulted in a loss of function.
A P P E N D I X

A.1 S U P P L E M E N T A R Y E X P E R I M E N T S

Experimental data additional to the results section.

Figure 54: Emission spectra of titrating Cy5 labeled AvrBs3 (B) to 5’-end Cy3 labeled specific DNA (I)

Example of emission spectra. Proteins were titrated in increasing concentrations (0, 0.99, 2, 4, 6, 10, 14, 18 [nM]) to 1 nM DNA (I). AvrBs3 (B)-Cy5 analyzed on DNA (I) (1nM).
Figure 55: Emission spectra of titrating Cy5 labeled AvrBs3 to 5’-end Cy3 labeled specific DNA (II)
Example of emission spectra. Proteins were titrated in increasing concentrations (0, 0.99, 2, 4, 6, 10, 14, 18 [nM]) to 1 nM DNA (I). (A) AvrBs3 (A)-Cy5 analyzed on DNA (II) (1nM). (B) AvrBs3 (B)-Cy5 analyzed on DNA (II) (1nM)
A.1 supplementary experiments

Figure 56: Spectra of DNA-Cy3 to determined the correction factor $\beta$

The correction factor $\beta$ had to be defined to correct the crosstalk between Cy3 and Cy5. Donor fluorophore was excited at 520 nm (ex$^D$) and emission was detected in the donor channel and in the acceptor channel. The factor $\beta$ is characterized by the fluorescence signal $F_{DexDemA}^{D}$ (A) divided with the fluorescence signal $F_{DexDemD}^{D}$ (B).

DNA-Cy3: $F_{DexDemA}^{D}$
DNA-Cy3: $F_{DexDemD}^{D}$


[121] LaFountaine, J. S., Fathe, K., & Smyth, H. D. (2015). Delivery and therapeutic applications of gene editing technologies ZFNs,


RPGR-associated retinal degeneration. *Human Molecular Genetics.* (Cited on pages 20, 21, and 107.)


“Es gibt Reichtümer, an denen man zugrunde geht, wenn man sie nicht mit anderen teilen kann.”

— Momo oder Die seltsame Geschichte von den Zeit-Dieben und von dem Kind, das den Menschen die gestohlene Zeit zurückbrachte von Michael Ende

ACKNOWLEDGMENTS

• I would like to thank Prof. Dr. Peter Friedhoff for becoming my first referee. Especially, I like to thank him for his enthusiasm about science, which steadily encouraged me to improve.

• I also like to thank Prof. Dr. Alexander Brehm for becoming my second referee and for his support as mine IRTG supervisor.

• In remembrance of Prof. Dr. Alfred Pingoud I like to thank him for accepting me in the IRTG and giving me the opportunity to be a part of the biochemistry institute at the JLU Gießen. In the short time I got to know him, he taught me a lot about science and about life.

• My thank also go to Dr. Wolfgang Wende for his kind support.

• I like to thank Prof. Dr. Dr. Knut Stieger for being a outstanding cooperation partner.

• I like to thank Dr. Mert Yanik for his practical advises at the beginning of my thesis, but also for being an excellent cooperation partner. I will keep in mind his big smile and hope all the best for his future.

• I also like to thank Professor Dr. Virginijus Siksnys for accepting to be my international IRTG advisor.

• My thank also go to Prof. Dr. Katja Sträßer for the smoothly association of the groups.
My thank go to all my coworker and friends who accompany me during the last three years:

- I like to thank Dr. Lilia Gabsalilow for teaching me the basics in biochemistry technicals and for all the fun hours together in our lab. She was the best partner to play tricks on Andi, Dennis and Laki and a very good room mate on our journeys.

- Also I like to thank Sabrina Stiehler for helping me in all imaginable situations...I will miss our everyday talks!

- I would like to thank Dr. Fabian Bietz for all his support in the lab and his patient to explain to me the L\LaTeX\ system. Particularly I enjoined to discuss with him about the coloration of our figures and about the most important question “what will be cooked tonight”.

- My thank also go to Julia Gotthard for the lovely trip to Hannover, for all the discussions during lunch time and for her support in the lab.

- I like to thank Dr. Maximilian Reuter for always providing support in the lab and I hope we will still meet up for that bottle of wine.

- I like to thank Dr. Anja Drescher for organization of the unforgettable trips with the IRTG (Moscow, St. Petersburg, and Rauischholzhausen). This unique time I will always keep in mind. I also like to thank Anja for her cheerful attitude and wish her all the best in Munich.

- I like to thank Ina and Karina for all support I got to fill in all possible documents.

- I like to thank Petra for being the best cleaning woman a lab can imagine.

- My thank go to Heike, Rashmi, Birte, Christoph and Pingping for completing our lovely group. I wish Pingping and her growing family all the best for the future.

- My thanks go to Dr. Bernhard Remes for his kind support and for the outstanding discussions during our coffee breaks.
• My thank go to Kristina Kovac‘ for the extraordinary discussion about life, while emptying a bottle of wine (or maybe two...).

• I like to thank my sister for her support, particular at end of the thesis.

• I like to thank Philipp for being my backup in all imaginable life situation. I like to thank him for his patient and for his support. I enjoy every second with you and hope we will have plenty of them in our common future.

• At the end I like to thank my parents for always supporting me and believing in me. Thanks for all possibilities in life I got from you and thanks that I can always trust you with my small and big problems which cross my way.

Giessen, January 2016

Laura Waltl